The processing of OVS and passive sentences in German monolingual and German-Italian simultaneous bilingual children

Doctoral dissertation submitted to the Faculty of Human Sciences at the University of Potsdam in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Cognitive Science

Chiara Boila

Supervisors:
Prof. Dr. Barbara Höhle, Universität Potsdam
Prof. Dr. Jacopo Torregrossa, Goethe Universität Frankfurt am Main

Reviewers:
Prof. Dr. Barbara Höhle, Universität Potsdam
Prof. Dr. Sarah Schimke, Ludwig-Maximilians-Universität München

Day of Submission: January 31, 2023
Thesis Defense: July 5, 2023

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-62972
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-629723

Acknowledgment

I am sitting here on my sofa and writing this last section. This means that I am done! A smile appears on my face. The journey was long and not always easy but with a lot of dedication I have reached this destination. Of course, I would not have been able to do that without the many people that supported me along the way. I would like to thank Professor Barbara Höhle for giving me the opportunity to embark on a PhD in the wonderful world of psycholinguistics and language acquisition. I had the opportunity to follow my research interests in the secure knowledge that Professor Barbara Höhle was there when I needed feedback. Professor Jacopo Torregrossa believed in me from the beginning, and I am glad to have had the possibility to be part of his team at the Goethe University of Frankfurt. Thank you for the inspiring discussions that we had about bilingualism and linguistics, and for your support and affinity. I am deeply grateful for the help that Tom Fritzsche gave me along the way. He was there for every question during the entire period of my PhD. From the beginning on he helped me to formulate my research questions and create the experimental designs. He was there until the end with a lot of time and patience during the statistical analyses. If R-programming is now a friend this is because of you. Thank you, Tom, for everything that you taught me during these years. I am grateful to Anna Walther for having recorded my German stimuli with great expertise and to my friend Dominique Hempe for drawing the amazing pictures of my experiments. I want to thank the Potsdam Graduate School (PoGS) for the financial support to participate in international conferences and summer schools. Special thanks also to the audiences of the EuroSLA 2018 (Münster, Germany), AMLaP 2018 (Berlin, Germany), the Cognitive Approaches in Language Learning and Bilingualism Workshop 2019 (Gothenburg, Sweden), the International Symposium on Monolingual and Bilingual Speech 2019 (Chania, Greece), the Summer School on Bilingualism and Multilingualism 2019 (Barcelona, Spain) and the Conference on Brain, Language and Learning 2019 (Siena, Italy). Your suggestions contributed improving my research and the quality of this dissertation.

And I want to thank all the people that I met at the BabyLAB in the years that I worked there. In particular, Vanessa, Peter, and Antonia, you were great colleagues and friends. Special thanks to Athina and Pia for helping me to digitalize around 160 questionnaires completed by parents and to evaluate and digitalize the standardized tests that the children performed. In total, we digitalized around 800 tests. This was a tremendous amount of work! Thank you for your essential help. I like to think of the BabyLAB Potsdam as the place where my scientific life started. My working years there were excellent. Thanks to Professor Barbara Höhle and Tom Fritzsche for this opportunity. I would also like to thank my colleagues at the Goethe University in Frankfurt, in particular, Chiara, Ioli and Andrea. It was a pleasure to spend time with you and discuss linguistics and life. I really enjoyed every moment.

Without families that participate in our studies with their children we could not do any language acquisition research at all. They are fundamental for our research. Thanks to them for the interest that they showed and for their availability. Additionally, I would like to thank the staff members of the German-Italian bilingual kindergartens in Berlin *Il Girasole*, *Il Piccolino*, *I Cuccioli*, *Rokokids* and *Brighella*, who gave me the possibility to test around 50 bilingual children. You offered me a room where I could test each child twice for around one hour each time. This was very generous and much appreciated. Meeting a lot of bilingual families, who showed time and patience, I could test 30 other bilingual children in their homes, again, twice for around one hour each time. And, counting the further 80 monolingual children that I tested at the BabyLAB (for approximately the same amount of time), I reached around 320 hours of only testing. I am glad that conducting this phase meant meeting a lot of nice people and wonderful bilingual children, who continually fascinated me by switching between their languages and even translating while at kindergarten.

Last, but not least I would like to thank Sebastian, my husband, who has supported me in every way during these years, especially in this last writing phase. Being and laughing with you and our daughter Emilia is the most beautiful thing of my life. Thank you to you both for making me happy.

Contents

Abstract

It is a common finding that preschoolers have difficulties in identifying who is doing what to whom in non-canonical sentences, such as (object-verb-subject) OVS and passive sentences in German. This dissertation investigates how German monolingual and German-Italian simultaneous bilingual children process German OVS sentences in Study 1 and German passives in Study 2. Offline data (i.e., accuracy data) and online data (i.e., eye-gaze and pupillometry data) were analyzed to explore whether children can assign thematic roles during sentence comprehension and processing. Executive functions, language-internal and -external factors were investigated as potential predictors for children's sentence comprehension and processing.

Throughout the literature, there are contradicting findings on the relation between language and executive functions. While some results show a bilingual cognitive advantage over monolingual speakers, others suggest there is no relationship between bilingualism and executive functions. If bilingual children possess more advanced executive function abilities than monolingual children, then this might also be reflected in a better performance on linguistic tasks. In the current studies monolingual and bilingual children were tested by means of two executive function tasks: the Flanker task and the task-switching paradigm. However, these findings showed no bilingual cognitive advantages and no better performance by bilingual children in the linguistic tasks. The performance was rather comparable between bilingual and monolingual children, or even better for the monolingual group. This may be due to cross-linguistic influences and language experience (i.e., language input and output). Italian was used because it does not syntactically overlap with the structure of German OVS sentences, and it only overlapped with one of the two types of sentence condition used for the passive study - considering the subject-(finite)verb alignment. The findings showed a better performance of bilingual children in the passive sentence structure that syntactically overlapped in the two languages, providing evidence for cross-linguistic influences.

Further factors for children's sentence comprehension were considered. The parents' education, the number of older siblings and language experience variables were derived from a language background questionnaire completed by parents. Scores of receptive vocabulary and grammar, visual and short-term memory and reasoning ability were measured by means of standardized tests. It was shown that higher German language experience by bilinguals correlates with better accuracy in German OVS sentences but not in passive sentences. Memory capacity had a positive effect on the comprehension of OVS and passive sentences in the bilingual group. Additionally, a role was played by executive function abilities in the comprehension of OVS sentences and not of passive sentences. It is suggested that executive function abilities might help children in the sentence comprehension task when the linguistic structures are not yet fully mastered.

Altogether, these findings show that bilinguals' poorer performance in the comprehension and processing of German OVS is mainly due to reduced language experience in German, and that the different performance of bilingual children with the two types of passives is mainly due to cross-linguistic influences.

Keywords: sentence processing, bilingualism, executive functions, language-internal and -external factors, eye-tracking, pupillometry.

List of Figures

List of Tables

List of Abbreviations

2L1	Simultaneous bilingual children
ACC	Accusative case
AoI	Area of Interest
AoO	Age of Onset
C°	Head of CP
CL	Clitic
CLLD	Clitic left dislocation
CP	Complementizer Phrase
CPM	Raven's Colored Progressive Matrices
DAT	Dative case
EF	Executive function
eL2	Early second language learners
ERP	Event-related potential
GEN	Genitive case
GSC	Global switch cost
I°	Head of IP
IP	Inflectional phrase
K-ABC	Kaufman Assessment Battery for Children
L1	First language
L2	Second language
LIFG	Left inferior frontal gyrus
LMM	Linear mixed-effect model
LSC	Local switch cost
MASC	Masculine nouns
ms	Milliseconds
NEUT	Neuter nouns

NOM	Nominative case
NP	Noun phrase
NP1	First noun phrase
NP2	Second noun phrase
OSV	Object-subject-verb
OVS	Object-verb-subject
PP	Prepositional phrase
PPVT	Peabody picture vocabulary test
RQ	Research question
RT	Reaction time
SD	Standard deviation
SE	Standard error
SES	Socio-economic status
SpecCP	Specifier position of CP
SVO	Subject-verb-object
TROG	Test for Reception of Grammar
V2	Verb-second
VP	Verb phrase

1 Introduction

The study of sentence processing is one of the major fields of psycholinguistic research. Its goal is to understand the mental mechanisms involved in sentence comprehension and processing, namely how listeners comprehend messages conveyed by spoken language or by written text. In language acquisition, one of the critical tasks in sentence comprehension and processing is learning how to identify who did what to whom: children need to learn how thematic roles (or theta-roles; θ-roles) are mapped onto syntactic functions and how syntactic functions are marked in the specific language. Sentences have elements that are minimally required and determined by the argument structure of the verb, which in turn determines the thematic roles of these elements. The relation between verbs and arguments is part of the lexical knowledge of the speakers (Chomsky, 1981; Fillmore, 1968; Gruber, 1976; Jackendoff, 1972). The thematic roles represent the interface between syntactic and semantic information, important for language processing and ambiguity resolution (MacDonald et al., 1994). Typically, in the case of transitive verbs (e.g., *meet*), the external argument-NP in the subject position is the agent of the event, whereas the internal argument-NP in the object position expresses the patient of the event. VP-internal NPs are obligatory with transitive and ditransitive verbs (e.g., *give*), but not with intransitive verbs (e.g., *run*). In case of a transitive verb, one VP-internal NP, a direct object, is required while a ditransitive verb requires two VP-internal NPs (or a NP and a PP), direct and indirect object. In contrast, intransitive verbs do not require a VP-internal NP. In order to interpret a sentence, each NP has to be paired to a thematic role that is in turn associated with a syntactic function (Haegeman, 1994). This means that children need to learn the theta grid of verbs in which the thematic roles are stored.

Some authors have suggested that the canonical linking between semantic and syntactic representations results from a hierarchical relation between the thematic roles of a verb (see for example Fillmore, 1968). It is proposed that the thematic hierarchy (e.g., Agent > Instrument > Theme/Patient according to Fillmore (1968), Agent > Experiencer > Theme/Patient in relation to psychological verbs according to Belletti and Rizzi (1988)) is the universal principle of subject selection by which the highest thematic role in the hierarchy is typically assigned to the sentence subject (see Scheepers et al., 2000 for the linking between thematic prominence and different types of verbs). Hence, the parser may develop a heuristic sentence processing by predicting the highest thematic role as subject and agent of the sentence. One interesting point is to analyze how the parser makes a choice when encountering an ambiguity during sentence processing. Temporary ambiguous sentences lead listeners to a ‘garden-path’ situation, in which the initial analysis of the sentence has to be revised in a reanalysis process. For example, in the well-known garden-path sentence from Bever' (1970) *The* *horse raced past the barn fell* the parser would initially misanalyse *raced* as the main verb of the sentence. It is only the last verb *fell* that resolves the ambiguity indicating that *raced* introduces a reduced relative clause rather than being the main verb. The parser may adhere to a principle of syntactic economy, in which the simplest syntactically permissible structure (determined by the number of nodes in the syntactic tree) is chosen ('Minimal Attachment Principle'; Franzier, 1978). According to Ferreira and Henderson (1998), the facility of the reanalysis process is affected by syntactic and thematic factors. Considering the sentence *Mary knew Susan would leave*, *Susan* is initially taken as the object of the sentence, because it is simpler than taking *Susan* as subject of the complement clause. A syntactic and thematic revision is required: *Susan* is not the object, but the subject of the complement clause and *Susan* is not the patient of *know* but the agent of *leave*.

In the past three decades, researchers have started to investigate children's language processing, providing, for example, information about how children interpret the sentences that they hear (Snedeker, 2013). Trueswell et al. (1999) studied the syntactic processing of garden-path sentences in children, such as *put the frog on the napkin in the box* employing a temporary PP-attachment ambiguity. Analyzing eye-gaze data and therefore providing information about the real-time processing of linguistic input, they proposed that listeners comprehend sentences incrementally by starting to interpret the linguistic input as soon as available, in order to understand which function is associated with which thematic role. Thus, the incremental processing account claims that children as well as adults process information immediately and that processing difficulties arise when a later cue in the sentence is in conflict with the preceding one (Macdonald et al., 1994; Trueswell & Tanenhaus, 1994; Trueswell & Gleitman, 2004). It has been hypothesized that while adults are able to revise their initial interpretation, children often fail (e.g., Trueswell & Tanenhaus, 1994).

It has been proposed that children's difficulties in revising their initial sentence interpretation might be related to their still-developing executive functions (EFs; also called executive control or cognitive control). EFs refer to a top-down mental process needed to concentrate and pay attention, in order to respond appropriately to salient stimuli in the environment. Responsible for inter alia attention, inhibition, shifting, problem solving and flexibility, EFs are a core component of self-control that regulates human cognition and action and enables humans to switch between different thoughts and actions (Barac & Bialystok, 2011; Diamond, 2013; Miyake & Friedman, 2012). Given that it is easier to continue doing a task than to change it, a great effort is required to switch the task (Diamond, 2013; Miyake & Friedman, 2012). Brain imaging studies have found activation localized to the left-inferior frontal gyrus (LIFG) under conditions of conflict, for example for incongruent trials in EF tasks and for condition of garden-pathing in sentence comprehension tasks (see Novick et al., 2005 for a review). It seems to be that the LIFG is the brain region involved in the control of suppressing an initially preferred interpretation and boosting an alternative. Being among the last neuroanatomical regions that mature, the LIFG may cause difficulties during children's reanalysis processing (Novick et al., 2005; Vuong & Martin, 2015). Hence, children's relying on early-arriving cues to sentence structure rather than late-arriving ones, especially in case of a conflict, may be due to their deficits in EF abilities, making the reanalysis of complex sentences difficult (Choi & Trueswell, 2010; Mazuka et al., 2009; Minai et al., 2012). Mature EFs are characterized by the ability to filter and suppress irrelevant information, thoughts, and actions in favor of relevant ones. Höhle et al. (2016) underlined the relevance of EFs when competition between different sentence interpretations arose. In their study children had to inhibit the canonical competitor - the default assignment of focus to the sentence object - to understand sentences with a focus to the sentence subject correctly.

Additionally, it has been proposed that in the domain-general executive functions is based the mechanism for managing attention in the bilingual mind (Bialystok, 1999). Bilinguals' experience in switching between two languages and different interlocutors might bring benefits for the development of EF abilities. However, in this last decade there has been a debate of whether the bilingual language experience modifies the development of EFs. While some results provide evidence that bilinguals have a cognitive advantage over their monolingual peers (see Bialystok, 2021 for a review), there are other findings which suggest no relationship between bilingualism and EFs (e.g., Paap et al., 2016). Nevertheless, if bilingual children possess more advanced EF abilities than monolingual children, this might be reflected in their more accurate performance in complex linguistic tasks (see Ostadghafour & Bialystok, 2021).

Beyond that, little has been investigated about how children acquire two languages in parallel. The majority of studies of language processing in bilinguals have focused on second language learners (e.g., Clahsen & Felser, 2006; Lago et al., 2021), and research on child bilingualism has typically used production and comprehension experiments, so that little is known about the real-time processing in bilingual children. To my knowledge, the processing of non-canonical sentences by bilingual children was only recently investigated in two dissertations: Cristante (2016) and Pontikas (2019). Cristante (2016) examined whether 7-year-old early second language learners of German (initial exposure to German between 3 and 4 years of age) with Turkish as L1 showed differences to German monolingual children during the comprehension and processing of German OVS and passive sentences. Her findings showed that the performance of bilingual and monolingual children was comparable during their comprehension of passive sentences, even though eye-gaze data revealed that bilinguals were slower than monolinguals in recognizing the voice cue to passive constructions. The comprehension and processing of OVS sentences was better in monolingual than in bilingual children. Pontikas (2019) investigated the comprehension and processing of object-which questions in English, by testing a monolingual and a bilingual group aged from 7 to 11. These bilingual children had different linguistic backgrounds. His findings showed in general similar accuracy rates in comprehension in both groups, but slower processing by bilingual children in comparison to monolinguals. For this dissertation I tested children that grew up bilingual from birth, namely simultaneous bilingual children. It may be that this group show similar processing costs with their monolingual counterparts. Other factors, such as EFs, language-internal and -external factors have also been considered to investigate the potential differences between monolingual and bilingual children during the comprehension and processing of German OVS and passive sentences.

1.1 The present dissertation

The studies included in this dissertation set out to contribute to the understanding of language processing mechanisms in monolingual and simultaneous bilingual children. Online methods (eye-gaze paradigms and pupil responses) were used with the main aim of providing new insights into monolingual and simultaneous bilingual children's language processing. OVS sentences were investigated in German monolingual and German-Italian bilingual 5;8-year-olds while passive sentences were investigated in German monolingual and German-Italian bilingual 4;6-year-olds. In German OVS sentences (e.g., *Den Hund schiebt das Schaf* – [the dog]$_{\text{MASC.ACC}}$ pushes [the sheep]$_{\text{NEUT.NOM/ACC}}$ – 'the sheep pushes the dog') the parser - driven by a subject-/agent-first bias - may initially take the object/patient of the sentence to be the subject and the agent. In German passive sentences (e.g., *Der Panther wurde gestern von dem Nilpferd gefangen* – [the panther]$_{\text{NOM}}$ was yesterday by [the hippopotamus]$_{\text{DAT}}$ caught – 'the panther was caught by the hippopotamus yesterday') the parser may initially take the subject of the sentence to be also the agent whilst in passive sentences the subject represents the patient. To correctly assign thematic roles in OVS sentences the parser needs to know German case-marking cues and in passive sentences, the parser needs to know the auxiliary voice cue and the optional prepositional *by*-phrase.

Sentence processing in monolingual and bilingual children might be influenced by the same factors (e.g., parents' education) or it might show different developmental patterns. The development of EFs, language-internal (e.g., age of acquisition of a language) and -external factors (e.g., language input in both languages) may play a role. Until now several studies have investigated possible outcomes of the interaction of the two languages in bilingual children mainly using spontaneous speech data (see Müller, 2017 for a recent review). Only few studies have focused on the morpho-syntactic processing of bilingual children and mostly in the area of grammatical gender agreement between determiners and nouns (Lemmerth & Hopp, 2019; Lew-Williams, 2017). Thus, the overarching research questions of this dissertation are:

Which factors influence monolingual and bilingual children's sentence comprehension and processing? And, what are the differences between monolingual and bilingual children's performance?

To answer these questions, offline (accuracy data during a sentence-picture verification task) and online measures (eye-gaze paradigms and pupil responses) were used. Offline comprehension tasks are normally based on behavioral responses such as manual pointing and/or verbal response after the language input ends. Online measures evaluate children's real-time sentence processing. While children listen to a word or a sentence that refers to a visual scene (for example pictures or a video) shown on a screen, gaze directions and pupil changes are recorded by devices like eye-tracking systems. Looks to the target should increase upon hearing the relevant linguistic cue. Each eye-movement takes around 200 ms and each saccadic eye-movement is preceded by a shift of attention. The listeners' eyes fixate on the element referred to in what they hear during the period of sentence processing. Thus, eye responses are closely time-locked to the input (Just & Carpenter, 1980; Matin et al., 1993; Van Engen & McLaughlin, 2018). Pupillometry has been employed as a psychophysiological index of cognitive load for multiple cognitive functions, including memory and language processing (Beatty, 1982; Kahneman & Beatty, 1966). Changes in pupil size may be an indicator of cognitive effort: larger dilation might indicate a processing effort (see e.g., van Engen & McLaughlin, 2018). Task-evoked pupillary responses are measured by comparing pupil dilation during an experimental task to a pupil size baseline measured immediately prior to task onset, normally between 250 and 500 ms (Beatty & Lucero-Wagoner, 2000; van Rij et al., 2019). As for eye-gaze data, pupil dilation is measured in response to a momentary event, such as an auditory stimulus presented in a listening task (e.g., Kret & Sjak-Shie, 2018; Zekveld et al., 2018). The task-evoked pupil response occurs spontaneously during stimulus presentation (Lum et al., 2017) and the pupil reaches its maximum size between one and two seconds after the presentation of a stimulus (Just and Carpenter, 1993; Piquado et al., 2010).

In addition to these measurements, the linguistic and cognitive profiles of children were assessed with standardized tests. Children's vocabulary and grammar knowledge, non-verbal reasoning ability and memory capacity were tested. These measures account for potential differences occurring within and between the bilingual and monolingual group during sentence comprehension and processing. Data from EF tasks, namely the Flanker task (Höhle et al., 2016; Rueda et al., 2004) measuring inhibition abilities, and the task-switching paradigm (Jersild, 1927; Wiseheart et al., 2016) measuring switching attention abilities, were collected with two aims: first, to explore whether the data supports the hypothesis of the bilingual cognitive advantage and second, to investigate whether children's performance in EF tasks affects their accuracy in the sentence comprehension task. Data gathered from a language background questionnaire were used to investigate whether children's sentence comprehension was affected by their parents' education or the number of older siblings in school, and for bilingual children the language experience in German was also correlated to their sentence comprehension.

This dissertation is divided into three parts: 'theoretical background', 'empirical investigations' and 'conclusions'. The 'theoretical background' part in Chapter 2 introduces the relevant linguistic properties of German and Italian syntax, in particular word order, case marking and passive structures. Chapter 3 describes factors influencing bilingual first language acquisition. The monolingual and bilingual acquisition of German OVS and passive sentences are reported in Chapter 4. The 'empirical investigations' part includes the two studies conducted: Chapter 5 describes the study regarding the comprehension and processing of German OVS sentences; monolingual and bilingual children's performance was analyzed, and results are reported. Chapter 6 reports the investigations conducted to analyze the comprehension and processing of passive sentences by children. Here, data are analyzed and findings discussed. The 'conclusions' part summarizes the results obtained and links them to possible factors that affect the comprehension and processing of German OVS and passive sentences in monolingual and bilingual children.

Part I Theoretical Background

2 Word order, case marking and passive constructions in German and Italian

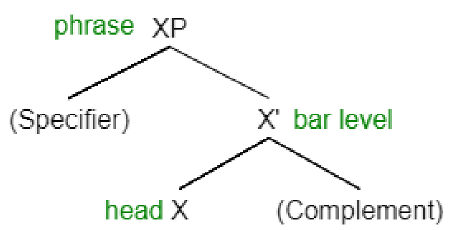
In this chapter I lay out the morphosyntactic properties of German and Italian which are necessary to understand the empirical investigations of this dissertation. When children acquire a language, they need to figure out the grammatical properties of their language, or languages in case of multilingualism. The word order properties of German and Italian are outlined first. Canonical main clauses share the same superficial word order in Italian and German but differ from each other in their underlying structure. The verb-second (i.e., V2) properties of German allow, for example, for a structure that is not possible in Italian, namely the fronting of an adverbial constituent to the sentence initial position and the placement of the subject after the finite verb in the third sentence position. Strongly connected with the word order properties is the case-marking system. In particular, the inflection in German determiners in the accusative case and object clitics in Italian are described. Lastly, the structure of long eventive passive sentences in German and Italian is presented.

2.1 Word order and word order variation

Within the Principle and Parameters Theory (Chomsky, 1995) of generative grammar in the context of the Government and Binding Theory (e.g., Chomsky, 1981, 1986), sentence structure is described as consisting of three phrases: the Verb Phrase (VP), the Inflectional Phrase (IP) and the Complementizer Phrase (CP), namely the C-domain, in which the IP is the complement of C° and the VP the complement of I°. Sentence structure is derived by the X-bar schema (Figure 1), according to which every language is organized by the same principles: every phrase has a head that may contain other phrases in the complement and specifier (or adjunct) position.

Figure 1

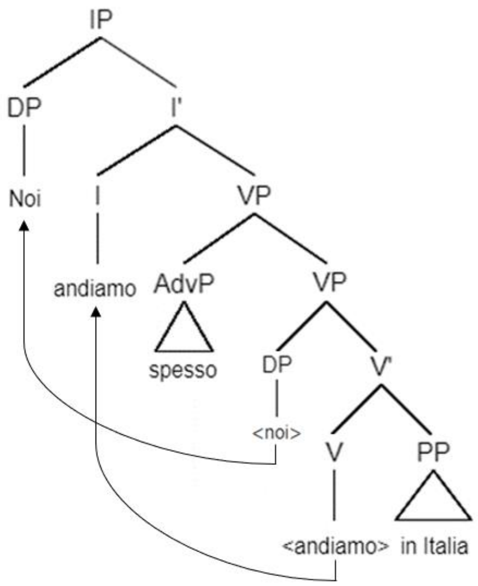
X-bar phrase structure



Languages differ in setting the head parameter (i.e., whether a complement precedes or follows a head) and movement operations (often responsible for word order phenomena). Considering, for example, verb movement and assuming a sentence structure within the Principle and Parameters framework the finite verb in Italian moves to I° (see Figure 2 and sentence (2.1)) whilst in German, a V2 language, it moves to C°.

Figure 2

Syntactic structure for the Italian sentence *Noi andiamo spesso in Italia*

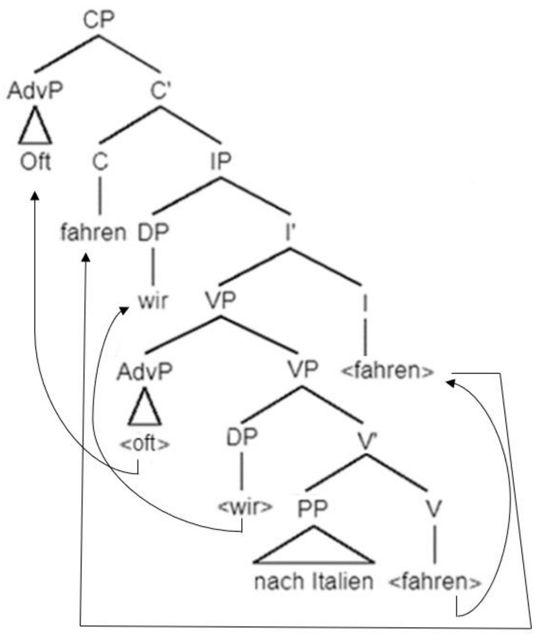


(2.1) *Noi andiamo spesso in Italia*
We go often to Italy
'We often go to Italy'.

The German sentence in (2.2) – corresponding syntax tree in Figure 3 – has the finite verb in C°. The position preceding the auxiliary or the finite verb in main clauses, the specifier position of CP (i.e., SpecCP), may be filled by only one other constituent (e.g., subject, object, adverbs, subordinate clause). Fronting constituents like adverbs or objects in German leads to sentence structures, in which the subject is in a postverbal position (see e.g., Vikner, 1994).

Figure 3

Syntactic structure for the German sentence *Oft fahren wir nach Italien*



(2.2) *Oft fahren wir nach Italien*
Often go we to Italy
'We often go to Italy'.

Word order variation in German usually involves the prefield, that corresponds to SpecCP, and the middlefield (Grewendorf, 2002). The middlefield in main clauses includes the area after the finite verb (V2) and is defined by one or more constituents.

Word order variation is influenced by information structure, in particular topic and focus. The preference of whether to have a subject or an adverb as first constituent of a sentence may depend on topicality: the subject is often the topic of the sentence and adverbs may be considered as frame-setting topics (e.g., Jacobs, 2001; Speyer, 2007). The object can occupy the prefield when it is a topic (i.e., topicalizations), a focus or a discourse-linked element (e.g., Frey, 2006; Sauermann, 2015; Speyer, 2007). Word order in the middlefield may result from scrambling, namely the positioning of an argument with respect to an adjunct (e.g., the direct object before or after the adverb) or to another argument (e.g., the direct object before or after the indirect object), (see for example G. Müller, 1999). Subjects and objects can be marked by case-marking cues on the determiner, an attribute adjective and in some instances on the noun itself.

Contrary to German, the SVO basic word order of Italian is also maintained when adjuncts are fronted, as for example the adverbial constituent in the next sentence (2.3).

(2.3) *Ieri Sebastian ha mangiato una mela*
Yesterday Sebastian has eaten an apple
'Sebastian ate an apple yesterday'.

Objects may be fronted by means of clitic left dislocations (CLLD) involving clitic resumption of the dislocated constituent (Belletti, 1999; Cruschina, 2010; Rizzi, 1997)[1]. A CLLD is shown in (2.4) in which the left dislocated topic, the direct object, is resumed by a clitic.

(2.4) *La mela, la mangia Sebastian*
The apple, $[\text{it}]_{\text{CL}}$ eats Sebastian
'Sebastian eats the apple'

The use of a clitic in left dislocation structures is necessary with direct objects and partitives (2.5), (Benincà, 1988; Cinque, 1990; Cruschina, 2010).

(2.5) *Sebastian, di mele, ne mangia molte*
Sebastian of apples $[\text{of them}]_{\text{CL}}$ eats a lot
'Sebastian eats a lot of apples'

Italian is a null subject language, which means that the subject may be dropped and that the statistically more frequent form in spoken corpora is (S)VO and O(S)V, not SVO (E. Bates, 1976).

To generate an object-before-subject linearization at the sentence surface, the movement of the corresponding object-NP to a position before the subject is required. In German this means that the object-NP moves to the already existing SpecCP (Scheepers et al., 2000). The object movement, from its canonical position as complement of the verb to the beginning of the sentence before the subject, is prevalently considered to be an A-bar movement (e.g., Frey, 2006; Friedmann et al., 2017).[2] On the other hand, it has been proposed that the CLLDed element in Italian is base-generated in its surface position and not derived by movement. Cinque (1990) argued inter alias that CLLD lacks the ability to enter into government chains which distinguishes it from its movement counterpart. Thus, despite

[1] Cinque (1990) distinguishes CLLD from left dislocation construction on different typological characteristics. For example, CLLD may involves several left dislocated elements as opposed to left dislocation; CLLD allows different XP (e.g., PP, VP) whereas left dislocation only allows NPs. For other typological differences see (Cinque, 1990).
[2] For a discussion on object scrambling and NP-movement as A-bar or A movement in Germanic linguistics see Broekhuis (2007).

differences in their syntactic structures, German and Italian OVS sentences both maintain the canonical linking between thematic roles and syntactic functions (subject/agent and object/patient) but, in both languages, the order of thematic roles is non-canonical since the patient (mapped to the NP1) appears before the agent.

Now, the question is which syntactic markers, or cues, do children use to map thematic roles onto syntactic functions. Children's comprehension of transitive sentences has received a great attention in the last forty years. E. Bates et al. (1984) found that animacy is the major cue to the agent in Italian 2-, 3-, 4-, and 5-year-olds as well as adults. The authors proposed that Italian children rely on animacy over word order cue probably because subjects/agents tend to be omitted. Additionally, the wide array of Italian word order variation might make the word order cue unreliable (MacWhinney et al., 1984). When animacy, prosody and subject-verb agreement are ambiguous or not available, Italian speakers have a strong bias to interpret noun-verb-noun sentences as SVO (Abbot-Smith & Serratrice, 2015). On the other hand, the most reliable cue for sentence interpretation in German is case marking. When case marking is ambiguous, adult speakers rely on word order, animacy and subject-verb agreement (MacWhinney et al., 1984).

2.2 Case marking

German has a rich case-marking system that distinguishes between four cases: nominative, accusative, dative and genitive. Case is marked on determiners, pronouns, adjectives and in some cases on the noun itself (e.g., accusative of weak masculine nouns, dative plural, genitive singular). The following table represents the case-marking paradigm of the NP in German.

Table 1

German case-marking systems

		Singular		Plural
	Masculine	**Feminine**	**Neuter**	**Masc/fem/neut**
Nominative	der Mann-Ø	die Frau-Ø	das Kind-Ø	die Menschen-Ø
Accusative	den Mann-Ø	die Frau-Ø	das Kind-Ø	die Menschen-Ø
Dative	dem Mann-Ø	der Frau-Ø	dem Kind-Ø	den Menschen-Ø
Genitive	des Mann-es	der Frau-Ø	des Kind-es	der Menschen-Ø
	the man	*the women*	*the child*	*the people*

A considerable amount of syncretism is present. Thus, concerning singular forms, masculine nouns are the only ones which have a distinct morphological marker on the determiner for all four cases. No overt difference occurs between the nominative and accusative cases for feminine and neuter nouns respectively, and between the dative and genitive cases for feminine nouns. The same form on determiners also occurs in the dative and genitive cases for neuter and masculine nouns respectively. Plural nouns are not marked in a way that distinguishes the nominative from the accusative case. The determiner of dative plural nouns overlaps with accusative masculine singular nouns while the determiner of genitive plural nouns overlaps with nominative masculine singular nouns and dative and genitive feminine singular nouns. In some instances, nouns also have a morphological marker. This holds for masculine and neuter nouns in the genitive case, as represented in Table 1. Additionally, some masculine nouns take an '-n' or '-en' in all cases except the nominative singular (the nominative case is considered to be the 'casus rectus', the neutral case with zero marking; Meisel, 1986). These nouns are the so-called 'weak' nouns and especially masculine nouns ending in '-e' take an additional '-n' (e.g., *der Junge*NOM., *den Jungen*ACC., *dem Jungen*DAT., *des Jungen*GEN., 'the boy'). Other masculine nouns, like 'the man' in Table 1, are the so-called 'strong' masculine nouns, namely they do not bear the inflection '-n' or '-en'.

Contrary to German, Italian does not mark case on determiners and/or nouns. Case marking is used for subject personal pronouns and clitics. Accusative and dative clitic pronouns in the 1st and 2nd person singular and plural are ambiguously case-marked, whereas accusative and dative clitic pronouns in the 3rd person singular are unambiguously case-marked. The 3rd person plural is unambiguously marked in the accusative case only (see Table 2)[3].

Table 2

Italian accusative and dative clitics

	Accusative		**Dative**	
	Masculine	**Feminine**	**Masculine**	**Feminine**
1 pers sg	*mi* 'me'	*mi* 'me'	*mi* 'me'	*mi* 'me'
2 pers sg	*ti* 'you'	*ti* 'you'	*ti* 'you'	*ti* 'you'
3 pers sg	*lo* 'him'	*la* 'her'	*gli* 'to him'	*le* 'to her'
1 pers pl	*ci* 'us'	*ci* 'us'	*ci* 'us'	*ci* 'us'
2 pers pl	*vi* 'you'	*vi* 'you'	*vi* 'you'	*vi* 'you'
3 pers pl	*li* 'them'	*le* 'them'	*gli* 'to them'	*gli* 'to them'

Note. For the use of *loro* 'to them' as weak dative plural pronoun see Cardinaletti and Starke (1999).

[3] Other clitic pronouns are locative (*ci/ce*, *vi*), reflexive (*si/se*) and partitive (*ne*) pronouns. There are no nominative clitics. For an introduction to Italian Clitics see Russi (2008).

Clitic object pronouns agree with the object in person and number and can be viewed as object agreement markers (MacWhinney et al., 1984).

To summarize, the object/patient in German OVS sentences can be indicated by means of case-marking cues on the determiner (and optionally on nouns in the case of weak nouns), whereas in Italian clitic left dislocations the object/patient is resumed by a sentence-internal clitic. This difference suggests that the performance of monolingual and bilingual children during sentence comprehension and processing might be different. For example, German-Italian bilingual children may look for a clitic as a cue for the identification of the object in German OVS sentences, as a cross-linguistic effect from Italian.

2.3 Passive construction

Active sentences and the corresponding passive sentences have the same content words and the same thematic roles are assigned to their arguments (Belletti & Guasti, 2015), as illustrated in (2.6) and (2.7):

(2.6) *The hippopotamus pushes the panther*

(2.7) *The panther is pushed by the hippopotamus*

Within the Principles and Parameters framework, the derivation of passive structures involves movement of the internal argument of the active construction (typically the patient) to the subject position maintaining the role of patient, while the external argument of the active structure (typically the agent) can be optionally realized as a prepositional phrase, namely the *by*-phrase in English. The passive voice, periphrastically expressed through an auxiliary in several languages, triggers an A-movement. Passive sentences lack the agentive external thematic role. This is due to the passive form of the auxiliary voice that fails to assign the accusative case. The patient of active sentences, the object, is the subject of passive sentences (see e.g., Carnie, 2007).

The auxiliary *sein* in German and the auxiliary *essere* 'to be' in Italian are associated with a stative reading (adjectival passives) while the German verb *werden* and the Italian *venire* allow for an eventive reading (verbal passives; see for German Abbot-Smith and Behrens, 2006; and for Italian Volpato et al., 2016). In German verbal passive structures, the finite verb is V2 and the *by*-phrase can be in the prefield (2.8) or in the middlefield (2.9).

(2.8) *Von dem Panther wurde das Nilpferd geschoben*

By the panther was the hippopotamus pushed

'The hippopotamus was pushed by the panther'.

(2.9) *Das Nilpferd wurde von dem Panther geschoben*
The hippopotamus was by the panther pushed
'The hippopotamus was pushed by the panther'.

Considering the S(finite)V alignment, passive structures overlap with each other across German and Italian (compare (2.9) and (2.10)), whereas the use of a constituent different from the subject in the prefield allows a sentence structure in German (e.g., (2.8) or adverb-auxiliary-subject) that is not possible in Italian.

(2.10) *L'ippopotamo venne spinto dalla pantera*
'The hippopotamus was pushed by the panther'.

To sum up, the subject/patient in long eventive passive sentences is indicated in German and Italian by the same cues: agreement with the auxiliary and the by-phrase. Although the lexical verb is final in German sentences, the linear order between the subject and the auxiliary may overlap between the two languages, if, for example, the German and the Italian subjects are in the preverbal position. In this case, monolingual and bilingual children are expected to behave similarly. However, German also allows for an adverb-auxiliary-subject order, with the subject/patient in the middlefeld. By contrast, the corresponding construction is ungrammatical in Italian. Due to the lack of overlap between the two languages, a different performance between bilinguals and monolinguals is expected in association with these structures, with bilinguals exhibiting slower processing and lower comprehension than monolingual children.

3 Factors influencing bilingual first language acquisition

Recent work on child L2 acquisition has explored several language-internal and language-external factors that affect bilingual language acquisition. Language-internal factors relate to language and cognitive development of the child itself. Internal factors are inter alias chronological age, age of onset and cognitive maturity (Paradis, 2011). It is important to consider these factors, or at least some of them, when developmental patterns of bilingual and monolingual language acquisition are compared. Language-external factors consider elements from the child's environment that may influence bilingual language acquisition. External factors include socio-economic status and parents' education as part of it or as a separate factor, number of siblings and/or family numbers, language input quantity and quality, language output (i.e., the language that the child is speaking), (Unsworth et al., 2011). These factors may also influence monolingual acquisition, but the main focus of this chapter lies on simultaneous bilingual acquisition (2L1), namely children acquiring two languages from birth, since this is the population under investigation in the studies of this dissertation. In this chapter I review previous studies investigating the role that language-internal and language-external factors may have on 2L1 acquisition. However, some data concerning language acquisition in children acquiring the second language before the age of 4 (namely early second language learners - eL2) called successive bilingual children (Rothweiler, 2006; Schulz & Grimm, 2019; Schwartz, 2004) are also reported. This was particularly done in order to understand the role of age of onset in bilingual language acquisition. Additionally, I have dedicated a separate section to cross-linguistic influence, that is: instances in which one language influences the other, and to the effects of bilingualism on cognitive development.

3.1 Language-internal factors

Age of onset (AoO) is the chronological age at which an individual begins to acquire a language and it is the most basic measure of exposure in child bilingualism. Several ages have been proposed at which the ability to learn another language native-like starts to decline. Children whose AoO is around 4 or later acquire a second language in a different way from children whose AoO is before this age (Meisel, 2009; Unsworth, 2016). The age of 7 is mostly considered a cut-off point for a critical (or sensitive) period for language acquisition, and therefore also for acquiring a further language native-like (Johnson & Newport, 1989; Meisel, 2011; Schulz & Grimm, 2019). Rothweiler (2006) examined the acquisition of V2, subject–verb agreement, and subordinate clauses in a longitudinal study with three L1 Turkish children acquiring German as second language (L2) aged from 3 to 5;7 years. The author observed that the children whose AoO was between 3 and 4 performed similarly to monolingual children, whereas children whose AoO was higher showed developmental patterns similar to adult L2 learners.

In addition, it has been found that the AoO relates with other factors like length of exposure, input and timing in the L1 acquisition, namely the chronological age at which a linguistic phenomenon is typically acquired. Unsworth et al. (2014) investigated the effect of AoO and the amount of input in the acquisition of grammatical gender by English-Greek and English-Dutch bilingual children aged between 4 and 9 years. The authors suggested a complex interplay between input, AoO and cross-linguistic influence. Considering the development of vocabulary, Unsworth (2016b) found that the amount of input was the only predictor for children's vocabulary score; the AoO did not play a role. Blom and Bosma (2016) argued that L2 vocabulary may grow faster in older than younger children, partly because they know more concepts and may profit from more cognitive resources, such as verbal short-term memory, attention span and declarative knowledge (for similar results on vocabulary see also Chondrogianni and Marinis (2011); Golberg et al. (2008); Snedeker and Geren (2007)). In the domain of syntax, they found that the length of exposure (rather than the AoO) had a positive effect on the correct use of inflection in 5- and 6-year-old bilingual Frisian-Dutch children.

Tsimpli (2014) proposed inter alias that the role of AoO in early bilingual development can be addressed when the timing in L1 acquisition is also considered. She explored whether the consequences of timing differences in monolingual development could be reflected in bilingual development. She differentiated between early, late and very late acquired phenomena in L1 acquisition. Early phenomena are core grammatical properties while late and very late phenomena involve semantics, pragmatics and non-verbal cognitive abilities. In line with Tsimpli (2014), Schulz and Grimm (2019) reported the subject-verb agreement as a grammatical property acquired early in German (in monolingual acquisition around the age of 3), the acquisition of wh-questions as late acquired (age of mastery in monolingual acquisition: age of 6), and the acquisition of case-marking as very late acquired phenomenon (age of mastery in monolingual acquisition: older than 6). Interestingly, they found that 2L1 children performed like monolingual children in the acquisition of early acquired phenomena, that they performed between monolingual and eL2 children for late acquired phenomena and worse than monolingual but like eL2 children for very late acquired phenomena.

These studies suggest that effects as of AoO do not occur in isolation but often in relation with length of exposure, input and timing in L1 acquisition. According to the AoO at which children start to learn a L2 their performance in linguistic tasks might be more similar to that of monolingual children or to that of late L2 learners.

3.2 Language-external factors

3.2.1 Socio-economic status

Previous studies on monolingual children's language development have consistently demonstrated effects of the socio-economic status (SES) on the development of linguistic skills in children. The SES variable might comprise parental education, parental occupation and parental income (Kohn, 1963). It has been found that children growing up in poverty can lag behind their peers in language performance (Hart & Risley, 1995). In the review by Hoff et al. (2002) it is reported that mothers with higher SES talk more to their children than those with lower SES, thus providing a greater variety of words and higher syntactic complexity. The speech of mothers with higher SES frequently has the aim of eliciting conversation while the speech of mothers with lower SES often has the purpose of directing the children's behavior. Similarly, in assessing the vocabulary of young children using the MacArthur Communicative Development Inventory (CDI; Fenson et al., 1993), Arriaga et al. (1998) found that 80% of low-income children scored below the 50th percentile. Huttenlocher et al. (2002) used two criteria to assess the SES. First, they gathered socio-economic information from different neighborhoods in the area of Chicago, which was then used to select preschools from which families were recruited for the study. Second, the maternal level of education was shown on a scale from 1 to 5 (1: some high school, 2: high school graduate, 3: some post high school training, 4: college graduate, 5: some graduate school). Mothers with a level of between 1 and 3 were categorized as having a lower SES. Mothers on levels 4 and 5 were categorized as having a middle SES. The study showed that the SES was not related to the production of complex sentences in 4-year-olds but only to their comprehension of complex sentences containing more than one clause.

In bilingual studies it is not always clear to what extent SES factors, input factors and bilingualism per se are responsible for differences in language and cognitive performance between monolingual and bilingual children. These factors are often correlated and some studies have attempted to disentangle them. Armon-Lotem et al. (2011) found for example that high levels of SES influenced the overall linguistic growth in bilingual children, especially of vocabulary size. They measured the SES by parent's education in years and occupational status (professionals, skilled workers and semi-skilled workers). Chiat & Polišenská (2016) assessed the SES of monolingual and bilingual children based on their neighborhood status (mid-high SES: inner-London neighborhood; low SES: outer-London neighborhood). Again, the performance of bilingual children in a vocabulary test was lower than that of monolingual children, as well as higher in mid-high SES than in low SES children. Bilingual children with low SES proved to be particularly disadvantaged. Meir and Armon-Lotem (2017) divided a bilingual group into low SES and mid-high SES subgroups, based on maternal and paternal education in years. There were no differences between the performance of bilingual children with low SES and mid-high SES in their L2 language proficiency (Russian). However, the children with mid-high SES outperformed those with low SES in their L1 (Hebrew).

The effects of SES have also been studied in relation to children's cognitive performance. Calvo and Bialystok (2014) investigated whether monolingual and bilingual children coming from working and middleclass families performed differently in EF and language ability tasks. The bilingual children had different L1 languages with a minimum length of exposure to English of two years. The SES was determined by their mothers' education on a scale from 1 (less than high school degree) to 5 (graduate degree). Maternal education correlated with paternal education and income index. Children from working class families had a smaller receptive vocabulary in English and were less accurate in inhibition and working memory tasks than children from middle class families. The impact of SES was similar for both monolingual and bilingual children. Furthermore, bilingual children obtained lower scores in an English vocabulary test than their monolingual counterparts, but showed higher performance in the inhibition and working memory tasks. Gathercole et al. (2016) tested bilingual children in Wales on English and Welsh with receptive vocabulary and grammar tasks in both languages, as well as general cognitive tasks. The parents' education on a scale from 1 (primary education) to 5 (post-graduate education) was applied as an SES index. It was found that the SES was a strong predictor of children's performance in vocabulary and grammar tasks in both languages and a weaker predictor on the cognitive tasks. Haman et al. (2017) focused on the development of vocabulary of bilingual children in their L1. Even though no differences occurred between monolingual and bilingual children in terms of SES (measured in terms of maternal education in years) and performance in cognitive ability tasks (non-verbal IQ and short-term memory), bilingual children showed a more limited vocabulary level in their L1.

To conclude, a number of previous studies has highlighted the role of SES for language and cognitive development in monolingual and bilingual children. Some studies have found that families' low SES negatively influenced the performance of monolingual and bilingual children in linguistic and cognitive tasks. However, it has also been found that this influence is dependent on the linguistic and cognitive tasks at hand, and on the L1 or L2 of bilingual children. Furthermore, the variety of methods used to assess the SES of families makes the comparison of results between studies difficult.

3.2.2 Language input and output

In recent decades the impact of input on bilingual children's language outcomes has been widely investigated. A range of factors influenced the role of input inside and outside the home. When children or older siblings start going to school, they may bring the school language home, thereby affecting the home language. The dominant language of the community may influence children’s language use (V. C. M. Gathercole, 2016). Jia and colleagues (2014) investigated Chinese–English and Korean–English bilingual children growing up in the United States. At a younger age, children commonly used Chinese and Korean with their siblings (60%-70% of time) while they mostly used English during their years at school (80% of time). Bridges and Hoff (2014) reported that Spanish-English bilingual children living in South Florida aged between 2 and 2;5 with school-aged siblings obtained a higher score with the CDI for English vocabulary and grammar compared to their peers without older siblings. This demonstrated that the school language spoken at home by older siblings influenced the linguistic development of their younger siblings. Other studies suggested that hearing a language from different speakers may be more supportive for language development than hearing a language from fewer speakers (Fisher et al., 2004; Place & Hoff, 2011; Richtsmeier et al., 2009).

The variation in quantity and quality of input may also have an impact on language development (e.g., Hoff & Core, 2013). Input quantity is probably the strongest factor to influence the language development of bilingual children (e.g., De Houwer, 2007; Thordardottir et al., 2006) and it varies according to differences in exposure at home, in the school and in the environment. ‘Richness’ measurements, namely how much time the child spends reading books in one language (or with parents reading aloud), or the time spent with other media or with extracurricular activities are considered as part of their environmental input (Jia & Fuse, 2007; Unsworth, 2016a). That is, input quantity refers to the amount of exposure available to the child and may be expressed in a relative or absolute amount of exposure (Unsworth et al., 2019). The relative amount of exposure is often calculated as the proportion of time that the child hears each language (at home and during daily activities), as reported by parents in questionnaires. The absolute amount of exposure typically analyzes the frequency of spoken words or syntactic constructions in spontaneous speech data addressed to the child. Using this measure, De Houwer (2014) found that bilingual children aged between 13 and 20 months received the same amount of exposure to maternal language input as monolingual children. Thordardottir (2014) reported that bilingual children with language exposure greater than 70% performed on a par with monolingual children in language production tasks, but that only language exposure of 40% was needed to perform similarly to their monolingual peers in language comprehension tasks.

Additionally, the type of language exposure that the child has, might influence language development. That is the so-called input quality, namely the type of exposure to which the child is exposed. Input-providers, who can be native or non-native speakers and may or may not speak a standard variety of a language, are factors that determine the type of exposure the child has (Cornips & Hulk, 2008; Unsworth et al., 2019, 2011). When parents speak a non-native language, they may provide less supportive input for language development than native speakers, who are normally more proficient in demonstrating more refined vocabulary, accurate phonology and morphosyntax (Hoff et al., 2014). Genesee et al. (2004) showed that input provided by immigrant parents with a high degree of English proficiency was a predictor for their children's language development. Similarly, Unsworth et al. (2019) found that the proficiency of non-native speakers was a predictor of bilingual children's performance in a receptive vocabulary task and in two productive morphosyntactic tasks. These findings suggest that the exposure to proficient non-native input might be more supportive for bilingual language development than previously thought.

Evidence illustrates that language input is often combined with language output, meaning how much time a person is speaking a language. Some studies found that language input and output are highly correlated (see for example Bedore et al., 2012; Unsworth, 2015), whereas other studies found that different language domains were differentially related to input and output (see e.g., Bohman et al., 2010). In the study of Bohman et al. (2010), for example, semantics was more related to input than output, whilst morphosyntax relied on both.

In summary, the school language that older children bring home may influence the language development of their younger siblings. The quantity and quality of input are factors that may considerably affect language acquisition and development. There is a need to determine which factors should always be selected for the calculation of the amount of input to assure that the findings between studies are comparable. Moreover, the type of input quality that children receive should also be reported. More research is needed in order to understand how these factors interact and influence each other in different linguistic domains.

3.3 Cross-linguistic influence

One of the main research questions involved in bilingual first language acquisition is: whether bilingual children develop their languages separately from one another or not. Are the words of the two languages stored independently or are they stored in a unique mental lexicon? Are the syntactic systems of the two languages acquired as one system or as two distinct systems? In the early nineteen-eighties, mixed utterances, meaning utterances or discourse containing features of both languages produced by bilingual children, led researchers to argue that children started their bilingual acquisition with a hybrid system merging features from both languages (e.g., Volterra & Taeschner, 1978). Language mixing was explained as a fusion of the two grammatical systems. This position is known as 'single system hypothesis' (or 'unitary language system hypothesis'). In the late nineteen-eighties/early nineties researchers provided evidence for the 'separate language hypothesis' (or 'autonomy hypothesis') arguing that bilingual children were able to distinguish between two linguistic systems from very early on (e.g., De Houwer, 1990; Genesee, 1989; Meisel, 1986). Today, the 'separate language hypothesis' is widely accepted, even though the two languages are in contact and may influence each other (Döpke, 1998; Hulk & van der Linden, 1996; Kupisch, 2007; N. Müller, 1998). The possible influence of one language on the other is referred to as cross-linguistic influence.

The ability to switch between languages within utterances or within a discourse is considered to be specific among bilingual speakers (N. Müller, 1998). In bilingualism, code-switching was defined as a specific characteristic of bilingual speakers in which they can switch between two languages, or mixing of elements from both languages within a single utterance (Meisel, 1989; Nicoladis & Genesee, 1997). Meisel (1989) described code-switching as a 'relief strategy' that children may use when the linguistic material is more accessible in one language than in the other. At the same time, Petersen (1988) formulated the 'dominant language hypothesis', according to which grammatical morphemes of the dominant language may be used with lexical morphemes of the dominant (e.g., *making* in English) and of the non-dominant language (e.g., *vaskING* – morpheme in Danish 'vask' for 'wash') while grammatical morphemes of the non-dominant language may only co-occur with lexical morphemes of the same non-dominant language (morpheme in Danish e.g., *mangler*: 'mangle' for 'lack' and '-r' as present tense marker). Along the same line, Genesee and colleagues (1995) found that French-English bilingual children tend to mix more when using the non-dominant rather than the dominant language. Lastly, at the beginning of the 21st century, Hulk and N. Müller (2000) and N. Müller and Hulk (2001) claimed that cross-linguistic influence depends more on language-internal factors (i.e., the properties of the grammatical domain in a language) rather than on language-external factors, such as language dominance. Some years later other researchers argued that both language-internal and -external factors have to be taken into account when predicting cross-linguistic influence (see e.g., Kupisch, 2007).

Furthermore, N. Müller and Hulk (2000) defined two criteria that motivate cross-linguistic influence. The first concerns a partial structural overlapping at the surface level of the two languages: if one language has two structures (e.g., German word order for simple clauses: SVO and SOV) and the other language presents only one of the two structures (e.g., Italian word order for simple clauses: SVO), then the structure common to both languages will be overused in the language allowing both structures (e.g., German SVO word order). In order words, when one of the two languages allows for more than one grammatical analysis and the other language for just one, bilingual children may show possible outcomes of cross-linguistic influence. The second criterion states that a cross-linguistic phenomenon occurs at the interface between grammatical modules, namely the syntax–discourse interface. Following Platzack (2001), who proposed that the C-domain may be responsible for causing problems in language acquisition, Hulk and N. Müller (2000) argued that in the C-domain, at the interface between syntax and pragmatics, the two languages may interact and thus cross-linguistic influence may occur. After all, the interaction between two modules of grammar is also hard to acquire for monolingual language acquisition (van der Linden & Sleeman, 2007). Sorace (2011) developed the 'interface hypothesis', stating that bilingual speakers may be less efficient than monolingual speakers in processing structures at the syntax–pragmatics interface because the syntactic processing is less automatic. Bilingual children may not be able to map universal strategies onto language-specific rules as quickly as monolingual children due to their confrontation with a much wider range of language-specific syntactic possibilities across the languages that they are exposed to (N. Müller & Hulk, 2001).

In addition, Paradis and Genesee (1996) defined three possible outcomes of cross-linguistic influence (defined in their paper as 'interdependence'): transfer, acceleration, and delay. A transfer occurs when a grammatical property of one language is incorporated into the other language. It may occur when the bilingual child reaches a more advanced level of syntactic complexity in one language than in the other. In other words, when the child is more dominant in one of his/her languages. A transfer may lead to types of grammatical mistakes that are not reported among monolingual children and that require a lot of time to disappear (N. Müller et al., 2007; Serratrice, 2013). Döpke (1998) investigated cross-linguistic influence between English and German in order to identify the mechanisms underlying verb placement in German-English bilingual children. Döpke (1998) showed two examples of transfer: first, German-English bilingual children produced SOV sentences in English, an error not found in English monolingual children (e.g., from a child aged 4: *can you that over bring* – 'can you bring that over?'); second, German-English bilingual children overgeneralized the (S)VO word order of English to German complex verb structures (e.g., from two 3-year-old children: *ich möchte tragen dich* – 'I want to carry you' and *du kann nicht kitzeln mich* – 'you can't tickle me'), a phenomenon not found in German monolingual children. In her seminal paper, N. Müller (1998) investigated the production of German subordinate clauses in monolingual and bilingual children, in which the L2 was English, French, or Italian. In German subordinate clauses the finite verb must be sentence final whereas in English, French or Italian it appears directly after the complementizers. This difference led bilingual children to produce target-deviant structures in German, in which the finite verb appears directly after the complementizers in subordinate clauses. N. Müller (1998) reported examples of target-deviant subordinate clauses from a French/German bilingual child at the age of 3;5: *dass das is ein baum* – 'that this is a tree'.

In literature on L2 acquisition a positive transfer is described as the effect that speeds up the acquisition process, while a negative transfer as the effect that slows down the acquisition process in comparison to monolingual children (N. Müller, 2016, 2017). This acceleration refers to the speed of acquisition of a certain property of grammar in a language that emerges earlier than would be the norm in monolingual acquisition. This is due to the children's linguistic knowledge achieved in the other language. Thus, the position of finite verbs in main clauses shares the same superficial SVO word order in German and Italian, even though the two languages differ in their underlying structures. German monolingual children have shown a target-deviant phase during the acquisition of the V2 rule, in which they place finite verbs at the end of the main clause instead of in the V2 position before the age of 4 (Clahsen et al., 1996). The placement of finite verbs in the V2 position in German main clauses is an example of acceleration in German-Italian bilingual children because their Italian may influence their German (N. Müller et al., 2007). Monolingual Italian children placed the finite verb from beginning on in the correct second position after the subject (Guasti, 1993). In the findings of N. Müller et al. (2002) a monolingual child produced the target-deviant position of finite verbs at the end of main clauses between the ages of 2;7 and 3;1 years whereas two German-Italian bilingual children did not show this pattern at the same age.

By contrast, delay indicates that the speed of acquisition of some grammatical phenomenon in one language emerges later than would be the norm in monolingual acquisition. The less complex analysis of a grammatical property of one language may be used by the bilingual child for both languages (N. Müller, 2017). An example of delay is subject omissions in Italian. German-Italian bilingual children produced more subjects in Italian than their monolingual counterparts, probably due to the influence of German, a non-null-subject language. This phenomenon was observed in both balanced and unbalanced bilingual children (measured in terms of mean length of utterance – MLU; Schmitz, 2007). Serratrice and Sorace (2002) produced similar findings, in which English-Italian bilingual children produced more subjects in Italian than their monolingual counterparts. English, like German, is a non-null-subject language. Serratrice et al. (2004) argued that cross-linguistic influence goes unidirectionally from the language with fewer pragmatic constraints (English) to the language regulated by pragmatically complex constraints (Italian) that only require the omission of pronominal subjects in certain circumstances (for example when the referent is already introduced in the context). English, the language with the more economical syntax–pragmatics interface system for subject pronouns, is assumed to influence the language with a more complex interface system, as in the case with Italian (see also Sorace, 2011). Thus, cross-linguistic influence will affect Italian, the more complex language, causing delay effects.

3.3.1 Cross-linguistic influence in language processing

Most of the cross-linguistic research done so far has taken language production into consideration. Since the early 2000s there has been the idea that differences in bilingual output compared to that of monolinguals, are due to processing demands. According to Grosjean (1998), bilinguals live in more than one language mode: bilingual speakers can be either in a monolingual language mode when they are speaking or listening to monolingual speakers in one language, or in a bilingual language mode when they are communicating with an individual who is sharing their languages and when code-switching takes place. When one language is activated, the other must be suppressed. Similarly, Hulk (2000) argued that for bilingual individuals in a monolingual context, one language is activated while the other is inhibited, even though influence from the other language cannot be completely suppressed and transfers might take place.

Works on cross-linguistic structural priming (or syntactic priming; Bock, 1986) attempted to show evidence for cross-linguistic influence driven by processing costs in bilingual speakers (Serratrice, 2013). Structural priming is explained as the facilitation of the processing of a syntactic structure by recent exposure to an identical or similar structure. It can be illustrated, for example, by reduced reading time for primed sentences in language comprehension studies, or the use of structures previously primed in language production studies. In bilingualism research, cross-linguistic structural priming is used to investigate syntactic representations of both languages in the bilingual mind (for a review see Van Gompel and Arai, 2018). The seminal paper of Hartsuiker et al. (2004) has inspired a lot of work in this field. They investigated whether syntactic information is shared in bilingual adult speakers, by applying cross-linguistic priming. The authors used structural priming from Spanish to English active and passive sentences to find evidence for the 'separate-syntax account' or the 'shared-syntax account'. According to the 'separate-syntax account', syntactic constructions that are superficially similar, but with some differences between the two languages (such as the presence or not of a preposition), are stored twice. The 'shared-syntax account' reduces redundancy because similar constructions in the two languages are represented only once, and differences between the two languages are stored as additional language-specific information. While it is possible that languages are distinctly represented, there is at least some degree of grammatical overlap between the two languages in which cross-linguistic influence occurs (Hartsuiker et al., 2004). In the case of sufficient similarity between English and Spanish active and passive structures, the shared-syntax account predicts cross-linguistic structural priming. In contrast, the separate-syntax account predicts no structural priming, namely no effect of the primed syntactic form on the target response. Spanish native speakers who had lived in the UK for two years on average were tested by Hartsuiker et al. (2004) in a production study using card sets. The experimenter spoke Spanish and the participants

English. Participants had to describe the pictures depicting actions of the card sets in a communication game with the experimenter. The findings showed that participants produced more English passive sentences after having heard Spanish passive sentences than Spanish active or intransitive sentences. These results supported the shared-syntax account.

Vasilyeva et al. (2010) studied whether bilingual children (mean age: 5;11) can extract a syntactic structure from a priming sentence and use it in a different language. They investigated priming from Spanish to English and from English to Spanish active and passive sentences in order to examine whether bilingual children had a shared representation of parallel structures, as proposed by Hartsuiker et al. (2004) for adults. Vasilyeva et al. (2010) found that children produced more English passive sentences after having heard Spanish passive sentences, but they did not find priming from English to Spanish. Some years later, Hsin et al. (2013) tested the hypothesis that all representations are shared between languages and that even constructions belonging to only one language should be available for their use in the other language. The authors used structural priming in which a grammatical utterance in one language (determiner-adjective-noun in English) should lead to the same utterance in the other language, despite being ungrammatical (determiner-adjective-noun in Spanish is ungrammatical; determiner-noun-adjective is grammatical). They found that 4 to 5-year-old Spanish-English bilingual children produced ungrammatical structures in Spanish after having heard the corresponding grammatical structure in English. It was argued that syntactic representations are shared across languages from an early age on.

The study of bilingual children online sentence processing has only started in the last decade and involves morphosyntactic processing focusing on grammatical violations in agreement of gender between nouns and adjectives (Lemmerth & Hopp, 2019; Lew-Williams, 2017). Previous work in this field had mostly used offline methods and reaction time (RT) paradigms, such as a self-paced listening task (see Blom & Vasić, 2011, for a study with Turkish-Dutch bilingual children; Chondrogianni et al., 2015, with Turkish-speaking bilingual children acquiring Dutch and English; Vasić et al., 2012, with Turkish-speaking bilingual children acquiring Dutch and Greek; Chondrogianni & Marinis, 2012, for the acquisition of tense morphology in English involving grammatical and ungrammatical sentences). As already mentioned in Chapter 1, the dissertation of Cristante(2016) and Pontikas (2019) were the first studies to investigate the processing of complex sentences in bilingual children. Both studies used eye-tracking paradigm and reported slower processing in bilingual children, compared to monolingual children. The children by Cristante (2016) were successive bilinguals (i.e., AoO: 3-4 years of age) and by Pontikas (2019) were simultaneous and successive (i.e., a group with heterogeneous AoO, range: 0-8 years of age and length of exposure: 2-10 years of age). The picture may be different when testing a homogeneous bilingual group with an AoO from birth indicating the same length of exposure as that of monolingual children.

In conclusion, the majority of data on cross-linguistic research was collected by means of behavioral methods and involved more language production than language comprehension experiments. It became evident that similar syntactic constructions in the two languages have a shared representation in the bilingual mind. Besides, online methods are necessary to tap the real-time properties of language processing.

3.4 Effects of bilingualism on cognitive development

Early research on the relation between bilingualism and cognition started in the 1920s. Cognitive ability was assessed based on performance in IQ tests as a measurement of intelligence. Saer (1923) tested English monolingual and English-Welsh bilingual children (aged from 7 to 11) by means of the Stanford-Binet intelligence test and described a mental confusion existing in bilingual children. The majority of studies in the first four decades of research reported that bilinguals suffered from a 'language handicap' and that this disadvantage could prevent the assessment of their intellectual abilities (Arsenian, 1967; Darcy, 1953, 1963; Macnamara, 1966). These findings led to statements regarding the negative effects of bilingualism (e.g., Hakuta & Diaz, 1985). Only some years later the findings of these early studies were treated as unreliable (see e.g., Cummins, 1976). According to Bialystok (2005) this early research paid no attention to factors such as the language of testing and the interpretation of results.

The belief that bilingualism had negative effects on children's cognitive development held until the publication of a study by Peal and Lambert (1962), often considered a turning point. They tested monolingual and bilingual 10-year-old children from the same school in Montreal. The actual knowledge of the two languages in children was thereby measured, using different tests, of which the results determined whether children were monolinguals or bilinguals, the bilingual sample having been more carefully selected than previously. The findings of this study showed that bilingual children outperformed monolingual children on verbal and non-verbal ability tests, especially when the task required mental flexibility.

In the early 2000s a new line of bilingualism research started, which focused on the effect of bilingualism on cognitive development, specifically on the set of EFs. The idea is that dual language experience in the first years of life has an impact on the cognitive development of the child. In order to manage two or more languages in daily practice, bilingual speakers have to suppress interferences from the non-target language while listening to or speaking the target language (Bialystok, 2005). The presence of a high level of EFs in bilingual populations is also suggested by the ease with which proficient bilinguals speak their languages and switch between their languages in appropriate contexts (Christoffels et al., 2013). This leads bilingual speakers to have a higher level of competence in using EFs, compared to monolingual speakers. Based on these findings, some studies have explained their findings as showing a bilingual cognitive advantage.

The first study to claim the existence of cognitive advantages in a bilingual population was Bain (1974). He tested 11-year-old monolingual and French-English bilingual children and found that the bilinguals performed better than their monolingual counterparts during problem-solving tasks. Bialystok and Majumder (1998) hypothesized that a balanced proficiency in both languages might be necessary for cognitive advantages to appear. They tested three groups of children in the third grade: a monolingual group and two bilingual groups with a differing degree of balanced bilingualism. A French-English bilingual group attending schools with formal instruction in French in an English-speaking metropolitan area was considered as the balanced group. A Bengali-English bilingual group was considered to be partially bilingual: these children attended schools with full-day instruction in English, while at home they heard both languages, English and Bengali. All the children indicated equal proficiency in a receptive vocabulary test for English, but the balanced bilingual children outperformed the other two groups when the tasks relied on selective attention. Likewise, Struys et al. (2015) found that 2L1 children had higher global accuracy rates than eL2 children on the Simon task, which measures domain-general inhibitory control abilities. De Cat et al. (2018) tried to identify a threshold for the emergence of a cognitive advantage in bilingual children. The authors calculated bilingual children's language experience as a continuous variable (measures of cumulative input and output) to identify how much bilingual language experience is necessary if they are to outperform monolingual children on the Simon task. They calculated a score (that represented 38% of bilingual children) above which bilinguals' performance was higher than that of the monolinguals. They suggested that the calculation of the score may be a start to identifying a criterion under which a bilingual population benefits from EF advantages. Bialystok (1999) investigated whether bilingual children showed a cognitive advantage over their monolingual peers using the Dimensional Change Card Sort task (Frye et al., 1995; Zelazo et al., 1996), a high-control task that requires selective attention to solve problems that are based on conflicting rules. Children were shown test cards depicting a red rabbit or a blue boat, and then they were given a set of red boat or blue rabbit cards, that had to be sorted first by color and then by shape, or vice versa. Hence, subjects needed to switch from one task to the other. Until the age of 4 or 5, children typically have difficulties in succeeding in the second phase of the task, namely when the sorting rule changes (Zelazo & Frye, 1997)[4]. The results indicated better performance by the bilingual children than the monolingual children in their ability to solve these problems.

In the last two decades research on EFs has become very popular in psycholinguistics and especially in research on bilingualism. A large number of studies have been published in which bilingual populations show cognitive advantages over monolinguals at different ages (for studies with infants see e.g., Kovács & Mehler, 2009; for studies with young children e.g., Bialystok & Barac, 2012; Bialystok & Martin, 2004; Martin-Rhee & Bialystok, 2008; for studies with young adults Costa et al., 2009, 2008; Prior & Macwhinney, 2010; Wiseheart et al., 2016; Yang et al., 2018; for studies with older adults e.g., Bialystok et al., 2004; Gold et al., 2013). Additionally, brain-imaging studies have revealed that the same brain areas (i.e., anterior cingulate cortex, the supplementary motor area and a cortico-subcortical network with the dorsal striatum) are involved in bilingual language control as in executive control processes, such as response inhibition and conflict monitoring (see Hervais-Adelman et al., 2015; Van Den Noort et al., 2019; for more information and neuroimaging results).

However, mixed findings and null results have also been found in this field (see Duñabeitia et al., 2014; Morton & Harper, 2007; Struys et al., 2018 for no evidence of bilingual advantage in children; see Kirk et al., 2014; Kousaie et al., 2014; Paap & Greenberg, 2013 for no differences between younger and older monolingual and bilingual adults). Methodological differences as well as other factors like SES, AoO and language proficiency may be responsible for the diverging findings reported in the corresponding literature (Morton & Harper, 2007). Van Den Noort et al. (2019) reviewed 46 published studies and found that research conducted from 2004 to 2012 reported strong evidence in favor of a bilingual advantage, whereas studies from 2013 to October 2018 reported more mixed findings and evidence against a bilingual advantage. Furthermore, de Bruin et al. (2015) claimed that any bilingual advantage may result from a publication bias favoring studies with positive results over studies with null results. The authors looked at around 100 conference abstracts from 1999 to 2012 on topics covering bilingualism, cognitive neuroscience and psycholinguistics. 63% of the studies that established a bilingual advantage were published by international scientific journals whereas only 36% of studies that reported mixed results or no differences were published. In a more recent study de

[4] The Dimensional Change Card Sort task bears a resemblance to another task used to measure shifting abilities, namely the task-switching paradigm (Jersild, 1927). Both tasks require participants to allocate their attention to a single task in the context of two potential options. Originally, there were differences between the two tasks: the Dimensional Change Card Sort task was a task with playing cards that had to be sorted in two boxes according to the color or shape cue, whereas the task-switching paradigm included the possibility of measurement RT and to use non-switch and switch trials (Cepeda et al., 2000). However, in recent years, modified versions of the Dimensional Change Card Sort task and the task-switching paradigm have been designed in a very similar way including different block types (color, shape, and mixed; see for example Wiseheart et al., 2016; Yang et al., 2018).

Bruin (2019) highlighted the importance of more detailed descriptions of bilingual samples, including the sociolinguistic context and how the acquisition of the two languages occurs, in order to identify more variables as to when a bilingual advantage might occur and when not. The importance of taking individual variability in research on bilingualism and cognition into account had also been pointed out (e.g., Bonfieni et al., 2020). Furthermore, Paap and his colleagues published a series of papers reporting a growing skepticism about the validity of claims for a bilingual advantage (see e.g., Paap & Greenberg, 2013; Paap et al., 2015, 2016).

Moreover, EFs have also been investigated in relation to complex sentence comprehension and processing. Höhle et al. (2016) found that better developed inhibitory skills were related to the ability of 4-year-old German monolingual children to interpret sentences correctly with the focus particle *only* in a non-canonical pre-subject position overriding the default-focus for objects in sentence final position. Minai et al. (2012) found evidence of a correlation between sentence processing by means of the universal quantifier *dono*-NP-*mo* (corresponding to the English every-NP) and children's EFs, measured as their task switching ability. Torregrossa et al. (2021) analyzed referring expressions used by bilingual children (mean age: 10;8 years), considering, among other factors, the effect of EFs. They found that children tended to produce full nouns in contexts where the use of null subjects or clitic pronouns (in object position) would have been more appropriate. The redundancy of children (when using referring expressions) correlated with low EFs. Teubner-Rhodes et al. (2016) tested Spanish-Catalan bilingual young adults and Spanish monolinguals with subject- and object-first cleft sentences and a cognitive-control task that included a no-conflict version, which omits the conflict trials, and a high-conflict version with conflict trials. Bilinguals outperformed monolinguals in the sentence comprehension task and only in the high-conflict version of the cognitive-control task suggesting that the bilingual advantage only emerges when the conflict monitoring demands are high. Similarly, Ostadghafour and Bialystok (2021) tested 4-year-old bilingual and monolingual children and found that bilingual children performed more accurately than monolingual children in a sentence-picture verification task.

Additionally, the efficiency of EFs required by different bilingual populations in different sociolinguistic contexts should also be taken into account. Mishra (2018) suggested that the control needed by bilingual speakers in a multilingual society that may encounter diverse types of interlocutors (in terms of spoken language and proficiency) may be different from the control needed by bilingual speakers usually meeting the same types of interlocutors in an almost monolingual country. In the first case, the control setting of bilingual speakers may need greater monitoring because of suddenly switching languages and proficiencies, whilst in the second case the control settings of bilingual speakers may be fixed without requiring high monitoring control.

It can be concluded that the bilingual advantage debate is still open. A central question in this debate is whether the tasks used to measure EFs activate the same control processes used by highly proficient bilinguals during language control (Friedman, 2019). The relation between EFs and language comprehension and processing also needs more research in order to be fully understood. Factors such as the age of the tested population and the tasks used to assess language comprehension and EF abilities seem to play an important role in the interpretation of results. Thus, research on bilingualism should continue in order to clarify the difficult questions in establishing relations between language and cognitive development.

4 German OVS and passive structures in L1 and 2L1 sentence comprehension

To acquire word order, children need to learn the conditions that may trigger word order variation. The purpose of this chapter is to provide a review of studies that have investigated how monolingual and bilingual children develop their understanding of German OVS and passive sentence structures, as well as how they weight different cues during sentence comprehension. Different accounts are presented in an attempt to explain children's non-adult like performance.

4.1 Children's comprehension and processing of OVS sentences

When processing a sentence, listeners follow cues such as word order and case marking to assign thematic roles. Trying to explain how children acquire the different linguistic cues of a language, such as word order and case marking, Macwhinney (1987) and Macwhinney and Bates (1989) developed the Competition model. Because different cues have different weights in different languages, the Competition model provides a way in which to establish the weight of a particular cue in a language. It deals with frequency in terms of cue availability (how often the cue occurs in the input) and cue reliability (how often the same cue points to the correct thematic role assignment). Thus, cue validity is a product of cue availability and cue reliability and varies across languages. In German transitive sentences, for example, the word order cue is widely available but not always reliable because German is a relatively free word order language with a complex case-marking system. The Competition model predicts that when two different cues are in conflict, the cue with higher validity will be used for thematic role assignment. However, this holds more for adults than young children, who seem to be more sensitive to cue availability than cue reliability (MacWhinney, 2007; McDonald, 1986).

According to Brandt et al. (2016) the word order cue in German points to correct thematic role in 79% of the sentences with two or more NPs whilst the case-marking cue to 100% of the sentences that contain at least one unambiguously marked NP. When case-marking is ambiguous, German adults seem to rely on further cues: the SVO word order (it is more often the case that the first noun is the subject/agent of the sentence), animacy and the subject-verb agreement (MacWhinney et al., 1984). Dittmar et al. (2008) demonstrated that 5-year-old German-speaking children rely more on word order than case marking, even though the case-marking cue has higher validity and should therefore be used for assigning thematic roles in German. It might be that word order is more transparent and easier to follow than case marking which has several syncretism in its system. The study by Dittmar et al. (2008) found that only children aged 7 performed above chance relying on case marking more than on word order cue, as adults did.

Abbot-Smith and Serratrice (2015) tested Italian-learning children aged 2;6, 3;6 and 4;6 on their comprehension of SO(pronoun)V sentences with novel verbs considering cue reliability, cue cost and cue conflict (e.g., *Il gatto lo baffa* – the cat itCL baff – 'the cat is baffing it') , SVO (*Il gatto baffa il cavallo* – the cat baff the horse – 'the cat is baffing the horse') and O(pronoun)VS (e.g., *Lo baffa il gatto* – itCL baff the cat – 'the cat is baffing it'). As mentioned in Section 2.2, case in Italian is assigned with subject pronouns and clitics. They found that SO(pronoun)V sentences were comprehended earlier than SVO(NP) sentences from the age of 2;6, demonstrating a correct comprehension of case-marked object pronouns in active sentences. O(pronoun)VS sentences were comprehended at chance level by all age groups. According to the authors, this was due to the predominant subjectless sentences in Italian, namely O(pronoun)V. Even though case is only available in 45% of Italian transitive sentences, it seems to be easier than in German, where case marking sometimes has the same forms with different grammatical functions. In fact, the accusative clitics marked on the third person singular used by Abbot-Smith and Serratrice (2015) show no syncretism within the clitic system, so that listeners can easily learn that when they occur before the auxiliary or finite verb, they can only be interpreted as the object of the sentence.

Additionally, Schipke et al. (2012) examined whether the neural system is able to detect the difference between nominative and accusative case-marking cues using an event-related potential (ERP) paradigm. German masculine nouns were used with unambiguous case marking on NP1 (i.e., *der* for nominative case and *den* for accusative case). 6-year-olds showed the same negative ERP response to object-initial accusative marked sentences as adults did, showing that the parser expected the nominative case on NP1. However, the positive ERP response to the nominative NP2 suggested that the 6-year-olds are able to process the initial accusative case marking on the NP1, but cannot use it for thematic role assignment, as also indicated by their chance level behavioral performance. At 4;6 years of age children showed no differences in the ERP signal between SVO and OVS sentences. 3-year-olds showed a positive ERP response to object-initial accusative marked sentences, thereby demonstrating that they were not completely insensitive to case-marking information. It was hypothesized that the positivity in the ERP signal showed a detection of acoustic differences between the nominative and accusative case cue on NP1 but no recognition of the underlying syntactic structure. In contrast, a recognition of these structures was shown in the negativity of the ERP signal by adults and 6-year-olds. A study by Biran and Ruigendijk (2015) tested 3-, 5- and 6-year-old German monolingual children using two types of OVS sentences: with a disambiguating cue on the NP1 (accusative masculine nouns) or on the NP2 (nominative masculine nouns). They performed a comprehension study using a sentence-picture verification task. The analysis showed no effect of age group and that the better performance on OVS sentences with a disambiguating cue on the NP1 was better than on the NP2.

It has also been argued that listeners may comprehend sentences incrementally by starting to interpret the linguistic input as soon as it is available, in order to understand which syntactic function is associated with which thematic role (see also Just & Carpenter, 1980). The finding that German adults start interpreting temporary ambiguous sentences with a subject-/agent-first bias has been reported in various studies (e.g., Bahlmann et al., 2007; Hanne et al., 2015; Knoeferle et al., 2005; Wendt et al., 2014). The difference between adults and children seems to be that adults normally revise their initial interpretation, whereas children do not start the revision process or fail throughout it. The incremental processing account (Macdonald et al., 1994; Trueswell & Tanenhaus, 1994; Trueswell & Gleitman, 2004) claims that children as well as adults process information immediately and that processing difficulties arise when a later cue in the sentence is in conflict with the preceding interpretation.

A further possible explanation for children's difficulties during complex sentence interpretation and/or processing is that their integration and storage of information during processing is less efficient than that of adults, mainly due to a short working memory span in children (S. E. Gathercole et al., 2004). Other authors followed the hypothesis that the presentation of non-canonical sentences in isolation may lead to higher processing costs in children. Indeed, a successive series of studies found that the processing difficulties of non-canonical structures are reduced when discourse information is included in the experimental design (see for OVS in German with adults: Burmester et al., 2014; Schumacher & Hung, 2012; for OVS in Finnish with adults: Kaiser & Trueswell, 2004). However, Sauermann (2016) did not find an influence of discourse information either in the comprehension and processing of German OVS sentences or in the comprehension of Russian OVS sentences (Sauermann & Gagarina, 2018).

Finally, a further explanation for OVS processing difficulties may be provided by the Relativized minimality approach (Rizzi, 1990)[5]. Sentences in which the object moves from its canonical position as complement of the verb to the beginning of the sentence, originating an A-bar movement, may cause problems for children because of the feature similarity of the moved object and the crossed subject (Friedmann et al., 2009, 2017). When the object or the subject is not a full NP but for example a pronoun, children may understand the structure better because pronominal may reduce intervention effects (Friedmann et al., 2017; Rizzi, 2013). In fact, Sauermann (2016) found that the accuracy of German children in the comprehension of OVS sentences increased when one referent was realized as subject or object pronoun compared to a full NP (for similar results see also Brandt et al.(2016)).

[5] Relativized minimality argues that a local relation cannot hold between two elements (X and Y) when a third element (Z), the potential candidate for the local relation, intervenes in between. This means that X must be related to Y, its trace, and that this relation fails when Z intervenes in the path between X and Y (…X…Z…Y…) matching the specification of X in the relevant morphosyntactic features triggering syntactic movements (intervention locality; Rizzi, 2013). The local syntactic process fails across elements that are structurally similar.

Moreover, it was found that in German child-directed speech the subject in OVS sentences is often realized as a pronoun (see corpus data in Dittmar et al. (2008) and Sauermann (2016)): frequency factors may enhance children's performance.

Turning now to the processing of German OVS sentences in a bilingual population, according to my knowledge, the dissertation of Cristante (2016) is the only work that has investigated these structures. She tested 7-year-old German monolingual and German-Turkish eL2 children with a sentence-picture verification task in an eye-tracking experiment. As in previous studies with German monolingual populations (Bahlmann et al., 2007; Biran & Ruigendijk, 2015; Wendt et al., 2014), she used two types of OVS sentences in which the disambiguating cue occurred on the NP1 (masculine accusative case marking) or on the NP2 (masculine nominative case marking). She created a between-subject design, in which one group of German monolingual and German-Turkish eL2 children was tested using SVO and OVS sentences with the disambiguating cue on the NP1 and another group was tested with SVO and OVS sentences with the disambiguating cue on the NP2. Accuracy data showed that all the children performed better in SVO than in OVS sentences of both types. Monolingual children's eye-gaze data demonstrated that they looked more to the target than to the distractor, while listening to SVO and OVS sentences with the disambiguating cue on the NP1. During the processing of OVS sentences, children's target fixations decreased just before performing the offline task. Turkish-German eL2 children looked more to the distractor than to the target during the processing of both types of OVS sentences. Regarding the processing of OVS sentences with a disambiguating cue on the NP2, the target looking by monolingual and bilingual children were comparable: children's target fixations were very low until the end of the sentence. They adopted an agent-first strategy that did not change after the presentation of the disambiguating cue.

Taken together, these studies have demonstrated that German-speaking children seem to rely more on word order than case-marking cue at least until the age of 7. The parser seems to be biased by a subject-/agent-first strategy and the acquisition of the German case-marking system seems to be a long process.

4.2 Children's comprehension and processing of passive sentences

In such languages as German and Italian, two cues are available in long passive sentences (which include the prepositional *by*-phrase): the auxiliary voice and the *by*-phrase. As mentioned in Section 2.3, the agent is mapped to the optional prepositional phrase and the patient of the action is mapped to the subject. Thus, the canonical alignment for agent-subject that is typical in actives does not hold for passives. If interpreted with the subject-/agent-first strategy (e.g., Meyer et al., 2012), passives require thematic reanalysis as soon as the cues that signal a passive structure are available. Once these cues have been identified, the thematic role of the NP1 needs to be changed from the agent to the patient. Ferreira (2003) tested psychology students in the US and found that they were more accurate and faster in identifying thematic roles in actives than passives. However, other studies with adults reported no differences between the accuracy data of passives and actives even though a greater activation of the left-inferior frontal gyrus was associated with passive-sentence processing (for neuroimaging studies see Feng et al., 2015, for Chinese and Mack et al., 2013, for English; for a bilingual population with English as L2 see Traxler et al., 2014).

Studies on the acquisition of passives have demonstrated that passives may be a complex structure for children. However, the age at which children master passive sentences seems to be language specific, and different types of passives also seem to influence children's performance. First studies in English have demonstrated that 4-year-olds have more difficulties in the comprehension and production of passives with non-actional verbs than in those with actional verbs (Maratsos et al., 1985) and in the comprehension and production of long passives than in that short passives lacking the *by*-phrase (Horgan, 1978). Horgan (1978) pointed out that short passives in English are ambiguous between stative (e.g., *the lamp was broken*) and eventive passives (e.g., *the lamp was broken* 'by the girl') while long passives may merely have an eventive reading. The role that stative and eventive passives play in children's acquisition was further examined by Borer and Wexler (1987), who proposed that children make more errors in eventive passives because they are not able to assign a thematic role to the moved subject NP. During the movement of the NP a trace is left in its original position, namely the VP. This trace forms a chain with the NP that is now in the specifier IP position. According to Borer and Wexler (1987), children have problems building these A-chains until the age of 5 or 6. On the other hand, in stative passives, as in the case of regular adjectival constructions, there is no argument movement involved and therefore children do not show difficulties. Nevertheless, Fox and Grodzinsky (1998) proposed that children's performance with stative and eventive passives is influenced rather by the use of short or long passives. They tested English children aged between 3 and 6 using a truth-value judgment task and only found children's performance at chance level in the condition that included non-actional verbs with long passive sentences (e.g., *the boy is seen by the horse*). Yet when children had to interpret short passives with non-actional verbs, their performance was above chance and their performance with actional verbs, independently of the use of short or long passives, was at ceiling. These results contrast with the account proposed by Borer and Wexler (1987). Fox and Grodzinsky (1998) suggested that these children's problem is related to the *by*-phrase and not to A-chains. The processing load increased by the transmission of the external thematic role to the *by*-phrase might exceed children's capacity and cause a break-down in the comprehension of passives.

The mixed picture that emerged from the acquisition literature of passive structures in English has also been documented in Italian. As in German, the Italian passive has two auxiliaries *essere* (in German *sein*) to express a stative reading and *venire* (in German *werden*) for an eventive reading. The interpretation of short passives with *essere* is like in ambiguous, as in English, whereby the word following the auxiliary can be interpreted either as an adjective or as a lexical verb (Belletti & Guasti, 2015). To investigate when Italian-speaking children master passive structures, Volpato et al. (2016) tested 3-, 4-, 5- and 6-year-olds on their comprehension and production of passive sentences using the auxiliary *venire* and *essere*, with actional and non-actional verbs and with or without the *by*-phrase. They applied a sentence-picture verification task to assess the children's comprehension and a picture description task to assess their production. They found in all groups that passives with actional verbs were better comprehended than passives with non-actional verbs. Differences between long and short passives were only found in the 3-year-old group, who also demonstrated a positive correlation between their accuracy during the comprehension of long passives and a forward and backward digit span task. The authors suggested that the memory capacity of 3-year-olds may overload during the processing of long passives. No differences were found between passives with *venire* and *essere*. Regarding production, all the children produced short and long passives without any significant differences, and they all preferred to produce eventive passives. These results are similar to the findings reported in one experiment by Manetti (2013). She tested 4-year-old Italian-speaking children by means of a syntactic priming paradigm, thereby expecting an increase in the production of passives due to the priming effect. Indeed, the production of passives was enhanced due to the priming effect, with a stronger effect for *venire* than for *essere*. However, in a task that allowed for more spontaneous answers in Manetti (2013), 4-year-olds produced more sentences with a pronominalized object (e.g., *La mucca lo lecca* – The cow [him]$_{CL}$ licks – 'The cow licks him') to answer patient-oriented questions than passives (e.g., *Il re viene leccato dalla mucca* – 'The king is licked by the cow'). This may suggest that passives still require complex syntactic operations. To summarize, 4-year-old Italian-speaking children seem to be able to use the grammatical functions and verbal morphology necessary to produce and comprehend passives with both auxiliaries correctly, even though the comprehension of stative passives is generally harder.

Previous studies on the acquisition of passives in German have reported that 3-year-olds were capable of producing passive sentences (Wittek & Tomasello, 2005). Regarding comprehension, different studies have demonstrated that, from the age of 5 years on, German-speaking children are above chance in interpreting passives. Accordingly, in a cross-linguistic study, Aschermann et al. (2004) tested 3-, 4-, 5- and 6-year-old German-speaking and English-speaking children on their comprehension of actives and long eventive passives. The German children were able to comprehend passive sentences like adults at the age of 5, while the English children were at least 6 years old before they could do so. Since the two languages are structurally similar in their passive formation, the authors argued that the differences found in comprehension were due to interactions with other aspects of the language grammar. Hence, the rigidity of the word order cue in English may lead children to rely on the NP1 as the subject of the sentence for a considerable time. In contrast, the relatively free word order in German may lead children to rely more on other reliable cues, such as case marking or subject-verb agreement. Armon-Lotem et al. (2016) tested 5-year-olds' ability to comprehend passives in 11 different languages, including German. Using a sentence-picture verification task, the authors compared short passives with active structures in a first experiment and long passives with active structures in a second experiment. The performance of children in all the languages was above chance for short passives. German children tested on short passives performed better than children tested on long passives. However, the level of success with long passives in German was also high, namely above 80%. High accuracy scores were also shown by Danish- and Dutch-speaking children, also V2 languages. Similarly to Aschermann et al. (2004), the authors argued that V2 languages may enhance the comprehension of the non-canonical assignment of thematic roles in passives because they allow for more flexibility in the word order. Dittmar et al. (2014) tested 2-, 3- and 4;5-year-old German-speaking children on their comprehension of passives using a forced-choice pointing paradigm with novel and familiar verbs. 2-year-olds preferred active sentences containing familiar verbs, following the agent-first strategy. 3-year-olds were at chance level with novel verbs and above chance with familiar verbs. Not until the age of 4;5 was the performance of children above chance with novel and familiar verbs, which suggests that even 4;5-year-olds possess the lexical and syntactic information necessary to understand passive constructions.

Only in the last decade have a few studies investigated the online processing of passive sentences in children. To my knowledge, three different languages were analyzed: Mandarin Chinese (Huang et al., 2013), English (Abbot-Smith et al., 2017; Huang et al., 2017) and German (Cristante, 2016; Ehrenhofer, 2018). The Mandarin Chinese-speaking 5-year-olds tested by Huang et al. (2013) demonstrated chance level performance in their comprehension of passives during an offline act-out task. By analyzing their online performance, the authors found that children were sensitive to the passive morphosyntactic cue *bei*. An important point of the study of Huang et al. (2013) was that passive sentences were interpreted more successfully when the passive cue *bei* appeared before the referential noun rather than after it. This means that when sentences contained a pronoun (i.e., the English equivalent *it*) in the first sentence position, *bei* in the second sentence position and the noun in the third, the sentences were easier for children to comprehend. The pronoun had the thematic role of patient and the noun that of agent. The authors suggested that the use of a pronoun may prevent children to assign the role of agent to a particular character and lead them to await the offset of the passive cue *bei* before they can assign a thematic role. This might have enhanced children's performance in passives when the noun appeared after the passive cue *bei*. Huang et al. (2017) also tested the same hypothesis in English-speaking children (mean of age: 4;11). A similar pattern to that of the study with Mandarin-Chinese children was found: English children performed better when passive sentences started with the pronoun *it* (e.g., *It is quickly eaten by the seal*) than with a full NP (e.g., *The seal is quickly eaten by it*). Following Huang et al. (2013, 2017), Ehrenhofer (2018) hypothesized that a voice cue prior to the subject allows the parser to avoid the agent-first bias. Consequently, children might be more successful at comprehending passives when the voice cue occurs before the subject. To test this hypothesis, she tested 5-year-old German-speaking children, because the V2 properties of German allow for passive structures with the voice cue before the subject (e.g., *Heute wurde die Robbe von ihm gefressen* - today was the seal by [it]DAT eaten - 'the seal was eaten by it today') and after it (e.g., *Die Robbe wurde heute von ihm gefressen* - the seal was today by [it]DAT eaten - 'the seal was eaten by it today'). However, act-out as well as eye-gaze results illustrated that children could correctly assign thematic roles in both active and passive sentences.

In addition, Abbot-Smith et al. (2017) conducted an eye-tracking experiment among 3-year-old English-speaking children using active and long passive sentences. The authors used novel verbs to investigate whether children show an agent-first bias (following Dittmar et al., 2014). Their results showed that children were able to distinguish between passive and active sentences and that for both structures the mean proportion of looks was on target. These online results were surprising, considering that previous offline studies among English children reported an age of mastering passives of between 5 and 6 years. The authors explained this contrast in terms of still-developing EFs that may be responsible for children's low performance in offline tasks. Cristante (2016) conducted an eye-tracking experiment on 7 and 10-year-old German monolingual and Turkish-German eL2 children (AoO in German was 3-4 years of age). To my knowledge, this is the only study that has tested German passive structures in bilingual children using an online method. Performing a sentence-picture verification task, German monolingual children showed accuracy scores almost at ceiling. Children's eye-tracking data did not display any differences to the adult group, thus demonstrating that German children from the age of 7 may process passive sentences correctly, relying more on morphosyntactic cues than on word order (see also Cristante & Schimke, 2020). Comparing the performance of monolingual and bilingual children, the authors found no differences in their offline performance, neither at the age of 7 nor at the age of 10. However, eye-tracking data showed that Turkish-German eL2 7-year-olds needed more time for the revision process: the target looking of German 7-year-olds was above chance, even before the onset of the *by*-phrase suggesting that the passive voice *wurde* was already a valid cue in recognizing passive structures. On the other side, the target looking of Turkish-German 7-year-olds was not above chance until the end of the sentence. The author explained these results in terms of length of exposure to German (3-4 years for eL2 children compared to 7 years for monolingual children), which may have played a role in children's sentence processing. Eye-tracking data of both 10-year-old groups proved that children started with an agent-first bias but, after the processing of *wurde*, they were able to revise their initial sentence comprehension. In this case, it was suggested that a length of exposure of 6-7 years to German was sufficient for eL2 children to perform similarly to their monolingual counterparts. Cristante (2016) suggested that the almost fully completed development of EFs by the age of 10 might have helped children to perform well in the offline task.

The online performance of the eL2 children tested by Cristante (2016) was comparable with the performance of Turkish-English eL2 children (AoO in English between 3 and 4 years old) in a study by Marinis (2007). He tested English monolingual and Turkish-English eL2 children aged between 6 and 9, using a self-paced listening paradigm combined with a sentence-picture verification task. In comparison to L1 children, eL2 children had longer RTs, but both groups showed the same pattern: children were capable of using morphological cues (*–ing*/*-ed*) to process actives and passives and to assign thematic roles. Thus, on comparing the looks to the target in Cristante (2016) and the RTs in the self-paced reading task in Marinis (2007), Turkish eL2 learners of German and English aged about 7 are able to recognize passive cues, even though a slower way than monolingual children. However, the offline results of Cristante (2016) showed a similar performance between monolingual and eL2 children while the results of Marinis (2007) demonstrated that eL2 children performed worse than monolingual children. Marinis (2007) argued that even when eL2 children were able to recognize the morphological cues for active and passive sentences, they rely more on word order in their final sentence interpretation.

Summarizing, it has been found that already at the age of 4 Italian-speaking children and at the age of 5 German-speaking children mastered passive sentences. The relatively rigid word order of certain languages, such as English, is considered as a factor that might hinder the correct assignment of thematic roles in passive constructions. Bilingual studies have shown that the length of exposure to a language can influence bilingual children's performance in their comprehension and processing of that language.

Part II Empirical Investigations

5 Study 1: German OVS sentence processing by L1 and 2L1 preschoolers

This study investigates whether German monolingual and German-Italian simultaneous bilingual 5;8-year-olds are able to use case-marking cues to assign thematic roles during the processing of German OVS transitive sentences with animate NPs that are case-marked on determiners. The canonical linking subject/agent and object/patient in OVS sentences is maintained, but the order of the syntactic functions and thematic roles is non-canonical: the object/patient appears before the subject/agent. So children have to inhibit the tendency to map the first argument into the subject, thus identifying the NP1 as object and patient of the sentence.

For this study, two types of OVS sentences were examined: OVS early and OVS late sentences. In OVS early sentences, the unambiguous accusative masculine case-marking cue appears on the NP1 (see 5.1). In OVS late sentences, the unambiguous nominative masculine case-marking cue appears on the NP2 - the NP1 is case ambiguous because in German, neuter nouns do not distinguish between nominative and accusative case (see 5.2).

(5.1) *Den Hund schiebt das Schaf*
[The dog]MASC.ACC pushes [the sheep]NEUT.NOM/ACC
'The sheep pushes the dog'.

(5.2) *Das Schaf schiebt der Hund*
[The sheep]NEUT.NOM/ACC pushes [the dog]MASC.NOM
'The sheep pushes the dog'.

It has been suggested that the length of the ambiguous region, or the position of the disambiguating cues, may influence sentence processing. The longer the ambiguous region (and the later the disambiguating cue) appears, the more difficult it is for the parser to abandon the initial sentence interpretation in favor of a later one (Bahlmann et al., 2007; Biran & Ruigendijk, 2015; Cristante, 2016). Consequently, the first research question (RQ) is defined as follows:

RQ1: Are OVS early sentences easier to comprehend and process than OVS late sentences for monolingual and simultaneous bilingual children?

It is hypothesized that OVS early sentences are easier than OVS late sentences, as well as in comprehending as in processing, for both groups. Assuming a subject-/agent-first bias, the unambiguously marked accusative NP1 might lead the parser to a correct assignment of thematic roles, whereas when the unambiguously marked nominative NP2 occurs in the sentence final position, the parser might be led through a revision process. It is hypothesized that the accuracy scores of children are higher for OVS early than for OVS late sentences. Concerning the online measures, the following predictions were made. Children should look more to the target picture while listening to OVS early than they do for OVS late sentences if OVS early sentences are easier to process. Regarding pupil responses, it is predicted that pupil dilation might be smaller during the processing of OVS early than of OVS late sentences, since larger pupil dilation may indicate a major effort during the processing of information (Beatty, 1982; Van Engen & McLaughlin, 2018).

Additionally, OVS early sentences were created using masculine weak and strong nouns. As mentioned in Section 2.2, weak nouns take the morphological cue '-n' or '-en' for the accusative form as a suffix on the noun (see 5.3), whereas strong masculine nouns remain uninflected (see 5.1).

(5.3) *Den Affen schiebt das Schaf*
[The monkey]MASC.ACC pushes [the sheep]NEUT.NOM/ACC
'The sheep pushes the monkey'.

Accordingly, the parser may obtain benefits from the additional cue during sentence processing. Thus, the next RQ is formulated as follows:

RQ2: Are OVS early sentences with weak nouns easier to comprehend than OVS early sentences with strong nouns for monolingual and simultaneous bilingual children?

It is hypothesized that monolingual and bilingual children may comprehend OVS early sentences with weak nouns better than OVS early sentences with strong nouns due to the further morphological cue as a suffix on the accusative noun. This means that higher accuracy scores, more looks to the target and smaller pupil dilation can be expected during the comprehension and processing of OVS early sentences with weak nouns.

Furthermore, as explained in Section 2.1, fronting the object in Italian triggers a clitic left dislocation (see 5.4). The clitic resumption (see 'lo' before the verb in (5.4)) leads to an OVS syntactic structure, which differs to that in German.

(5.4) *Il cane, lo spinge la pecora*
The dog, [it]CL.MASC.ACC pushes the sheep
'The sheep pushes the dog'.

Thus, the next RQ is:

RQ3: Do bilingual children show a different comprehension and processing of OVS sentences in comparison to monolingual children?

According to N. Müller and Hulk (2000), cross-linguistic influence can occur in the syntax-discourse interface. Syntactic processing among bilingual children might be less automatic than that among monolingual children because they have less knowledge of representations or less access to these representations (Sorace, 2011). The fact that German OVS sentences have a different syntactic structure than Italian OVS sentences and the assumption that both languages are active in bilingual speakers even when only one language is used (Dussias & Sagarra, 2007; Kroll, 2008; Kroll et al., 2012; Marian & Spivey, 2003), lead one to hypothesize that German-Italian bilingual children may show lower accuracy scores, less looks to the target and larger pupil dilation than monolingual children do.

Additionally, since diverging results have been found between offline and online measurements (e.g., Brandt-Kobele & Höhle, 2010; Höhle et al., 2016; Sekerina et al., 2004), the next RQ concerns the method and is formulated as follows:

RQ4: Does the performance of monolingual and bilingual children differ according to offline and online measures?

Some studies have shown that children displayed a more adult-like performance with non-canonical sentences when using online rather than offline measures (e.g., Höhle et al., 2016; Sekerina et al., 2004). One assumption is that the complexity of the offline tasks may influence children's responses by leading them to low accuracy scores (for example below chance level), which is not in line with the looks to the target. Children might look to the target picture and then give a wrong answer in the offline task, thus their language comprehension abilities could potentially be underestimated.

One proposal for children's low performance in offline measures is that they may be influenced by their still-developing EFs (see e.g., Choi & Trueswell, 2010; Cristante & Schimke, 2020; Höhle et al., 2016; Novick et al., 2013). To explore whether or not this is the case, two EF tasks were employed. Children's inhibition and switching ability were measured. Thus, the next RQ can be formulated:

RQ5: Are monolingual and bilingual children's EFs related to their performance in the comprehension of German OVS sentences?

A relation between accuracy scores on sentence comprehension and children's EF abilities is only expected during the comprehension of OVS (and not SVO) sentences, i.e., the complex sentence condition. Better inhibition and switching attention abilities might help children in their correct assignment of thematic role in OVS sentences, (whereas in SVO sentences no reassignment is required). Advanced EF abilities might produce high levels of performance in the linguistic task if the linguistic structure is difficult to master.

If EF abilities are a predictor for children's performance in the OVS sentence comprehension task, higher accuracy scores should correlate with the performance on the EF tasks. Since a large body of literature (see Section 3.4) has reported that bilingual children show a cognitive advantage over their monolingual counterparts, they may outperform the monolingual group in the EF tasks and, accordingly, in the sentence comprehension task. However, if children's performance on the EF task does not predict the comprehension of OVS sentences, bilingual children's performance on the sentence comprehension task might be even less accurate than that of monolinguals, due to influences by other language-external factors (e.g., language input). This study considers the following language-external factors: language experience for bilingual children (i.e., the amount of input received by bilingual children in different situations), the parents' education and the number of school-aged older siblings present in the families of both bilingual and monolingual children. Consequently, the next RQ is:

RQ6: Do language-external factors predict monolingual and bilingual children's comprehension of German OVS sentences?

Bilingual children's accuracy scores in the comprehension of OVS sentences may positively correlate with their language experience in German. A higher level of education in their parents and a higher number of school-aged older siblings are assumed to predict the performance of monolingual and bilingual children in the OVS sentence comprehension task.

Finally, all the children were given a standardized test on receptive vocabulary and grammar, verbal and visual short-term memory and non-verbal reasoning ability. This was done in order to assure comparability between the groups on more general linguistic and cognitive tasks. Additionally, previous studies have found that monolingual children perform better than bilinguals in receptive vocabulary tasks (e.g., Bialystok et al., 2009), or that bilingual children outperform monolinguals in tasks assessing their memory capacity (e.g., Blom et al., 2014) or their reasoning skills (e.g., Tsimpli et al., 2020). It could be assumed that the comprehension of non-canonical sentences is influenced by children's general knowledge of language and non-verbal reasoning ability and that memory capacity might modulate children's sentence processing (e.g., Arosio & Giustolisi, 2019; Arosio et al., 2010).

Thus the last RQ is:

RQ7: Is the comprehension of German OVS sentences in monolingual and bilingual children influenced by their language knowledge, memory capacity and reasoning skills?

Higher accuracy scores in OVS comprehension are supposed to correlate with higher scores in all of these tests. A further assumption is that language-external factors, as well as more general linguistic and cognitive abilities, only affect the sentence type requiring more effort to be processed, namely OVS late sentences.

In the next sections of this chapter, I first describe the participants and the materials used and then the empirical findings obtained through quantitative data analyses. Finally, the discussion section summarizes and discusses the data.

5.1 Method

5.1.1 Participants

In total, 79 5-year-old children participated in this study, 40 monolingual German-learning children and 39 simultaneous bilingual German-Italian learning children. The selection criteria for bilingual children were exposure to both languages from birth and families following the one parent-one language approach. However, two children having only Italian input at home were included in the analysis because they obtained normal scores in a German vocabulary and grammar test standardized for a German monolingual population (with minimum *z*-score of -1, i.e., 16th percentile, and a maximum *z*-score of 2.3, i.e., 99th percentile). Furthermore, their inclusion in the analyses did not affect any of the results. From this sample, 8 bilingual children had to be excluded for the following reasons: childhood anxiety disorder (n=1), contacts with a third language at home or from a previous environment (n=5), selective mutism (n=1), no understanding of Italian (n=1). In total, four monolingual children were excluded from the analysis: three children did not complete the second test session and one did not reach the minimum score of 65% accuracy in the SVO condition of the sentence comprehension task. Additionally, in order to ensure the same sample size in both groups, children were matched by age. Accordingly, five other monolingual children were excluded from the analysis.

The final sample of this study consisted of 31 monolinguals (mean age: 5;8, range: 4;11-6;7; SD: 0;6, 17 girls) and 31 bilinguals (mean age: 5;8, range: 5;1-6;8, SD: 0;6, 17 girls). In bilingual families, the mothers were native speakers of Italian in 17 cases and native speakers of German in nine cases. Parents were considered as native speakers when their AoO was in that language at birth. In five families, mothers had also been raised bilingually, namely AoO from birth both in German and Italian. Three of them spoke Italian to their children, one spoke German and the other one both languages. In

13 families, the fathers spoke primarily Italian to the children and in 18 families they spoke prevalently German. Two bilingual children had no German native speakers in their families, as already mentioned. One of these two children went to a German monolingual kindergarten and the other to a German-Italian bilingual kindergarten. Of the 29 remaining bilingual children, 18 went to a monolingual kindergarten and 11 to a bilingual one.

According to parental report, these children had no indications of atypical development. All children were recruited from the Berlin/Potsdam area. Monolingual children were tested at the BabyLAB of the University of Potsdam and bilingual children were tested in the kindergartens or at home (except two bilingual children who were also tested at the BabyLAB of the University of Potsdam). Parents were reimbursed for their travel costs to the lab and the children each received a small book after completion of the experiment. This study was approved by the Ethics Committees of the University of Potsdam. Written informed consent was obtained from all parents.

5.1.2 Materials and Procedure

5.1.2.1 Language exposure and socio-economic profile

A language background questionnaire (see Appendix A.1) inspired by previously published questionnaires (BiLEC: Bilingual Language Experience Calculator from Unsworth, 2013; MAIN: Multilingual Assessment Instrument for Narratives from Gagarina et al., 2012; LEAP-Q: The Language Experience and Proficiency Questionnaire from Marian et al., 2007; modified version of the LEAP-Q from Souza et al., 2013; LSBQ: Language and Social Background Questionnaire from Anderson et al., 2018) was developed. This allowed inter alia to assess bilingual children's current patterns of language exposure and to calculate the cumulative quantity of input. This questionnaire included general questions regarding birth conditions and the child's development, language exposure from birth and in the following years: the language(s) spoken by caregivers with the child during an average week, as well as the language(s) spoken by the child with the parents, the amount of early literacy exposure for the two languages (time spent by parents reading out loud), the language in which the child watched television, listened to music, played videogames and practiced sport. Parents of monolingual children also completed a language background questionnaire, that was congruent with the bilingual one, except for index scores regarding language input and language output (see Appendix A.2). Parents were asked to state their level of education on a scale from one to five (1: elementary school, 2: 10th class / apprenticeship, 3: A level, 4: university degree, 5: PhD degree), and how many children lived at home, and their ages. The variables used are reported in Section 5.2.1.

5.1.2.2 Standard assessments

Linguistic profile

To assess the receptive vocabulary of the children, the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2007) was adopted. The monolingual children completed the German version of the PPVT (PPVT-4; Lenhard et al., 2015) while the bilingual children also completed the Italian version of the PPVT in addition to the German version (PPVT-R; Stella et al., 2000). In this task, children were asked to select one out of four pictures that matched a word produced by the experimenter following the standard protocol. The German and the Italian versions differ from each other in terms of visual presentation and dropout criteria. The German PPVT contains colored pictures, a total number of 228 items and a dropout criterion of four errors in a predefined set of 12 items. The Italian PPVT contains black and white pictures, 175 items in total and a dropout criterion of six errors in eight consecutive answers. A degree of balanced bilingualism (balanced vs. unbalanced) was calculated as the ratio of German PPVT score to Italian PPVT, in which perfect balance would produce a score of 1. All bilingual children obtained scores higher than 1, thereby indicating a vocabulary proficiency stronger in German than in Italian.

The Test for Reception of Grammar (TROG; Bishop, 2003) was used to assess the receptive grammar. The German version (TROG-D; Fox, 2009) was administered to both monolinguals and bilinguals. Bilinguals also performed the Italian version (TROG-2; Suraniti et al., 2009). Similar to the PPVT, children were asked to select one out of four pictures that matched a sentence presented by the experimenter. For both languages the test was aborted with four consecutive wrong blocks. A block was considered wrong when at least one answer out four was incorrect. The German TROG has a total of 84 items, in which the first 12 items are words and not sentences, while the Italian TROG has a total of 80 items, all of which are sentences. To calculate the degree of balanced bilingualism (again, balanced vs. unbalanced) the 12 word items in the German version were not considered. Children's raw scores obtained in the German TROG were then divided by 72 and the raw scores obtained in the Italian TROG by 80. Thereafter, I proceeded as for the PPVT, calculating the ratio of German TROG score to Italian TROG. Results from both linguistic tests are reported in Section 5.2.2. To compare groups, raw scores of children's performance in the German PPVT and in the German TROG were used as dependent variables.

Cognitive profile

As a measure of reasoning ability from a non-verbal perspective, a German version (Bulheller & Häcker, 2002) of the Raven's Colored Progressive Matrices (hereafter CPM; Raven, 2003) was administered.

The test consists of three sets with twelve items each. In each item a pattern with a missing part is presented. Six similarly shaped pieces were shown under the pattern with the missing part. The child was required to point to the piece that completed the pattern.

To assess memory capacity in monolinguals and bilinguals, the word order and number recall subtests from the German version (Melchers & Melchers, 2015) of the K-ABC II (Kaufman et al., 2003) were employed. In the word order subtest (hereafter visual memory), the experimenter read the names of common objects (e.g., house, bird, star). The strings ranged from one to six words. Each child had to touch a series of silhouettes of these objects illustrated on a sheet of paper, in the same order in which they were read out. After having reached the 18th item (the test contains 27 items), children were shown a series of small, colored squares. From this item onwards, upon hearing the names of the objects and before touching them, children were prompted to say the colors of the small squares for five seconds. This is the test part related with working memory. The test terminated when the child made three consecutive errors. In the number recall subtest (known as digit span and hereafter named digit span), the experimenter read a string of numbers, and the children were asked to repeat the string in the same order. The strings ranged from two to nine digits. This test has 21 items and it has to be aborted after three consecutive wrong answers. It is used to measure memory span. The results of all three cognitive tests are also reported in Section 5.2.2. To compare groups, the raw score of participants' performance was used as dependent variables for all three tests.

5.1.2.3 Sentence comprehension and processing

Design and materials

32 experimental items were created, of which each one described a scene with two animal characters: the agent and the patient. The grammatical gender of eight animal labels was neuter and that of 24 animal labels was masculine. Half the masculine nouns belonged to the weak declination type (i.e., masculine nouns that bear an inflection '-n' or '-en' in the accusative singular case) and the other half were strong nouns (i.e., masculine nouns that do not receive any overt inflections on the stem in the accusative case). The list of animal characters is reported in Table 3.

Table 3

Stimuli list OVS sentences

Neuter nouns	**Masculine weak nouns**	**Masculine strong nouns**
das Huhn ('hen')	*der Hase* ('rabbit')	*der Hund* ('dog')
das Küken ('chick')	*der Koalabär* ('koala')	*der Dino* ('dinosaur')
das Seepferdchen ('seahorse')	*der Bulle* ('bull')	*der Delfin* ('dolphin')
das Lama ('lama')	*der Affe* ('monkey')	*der Wolf* ('wolf')
das Krokodil ('crocodile')	*der Leopard* ('leopard')	*der Igel* ('hedgehog')
das Nashorn ('rhinoceros')	*der Pandabär* ('panda')	*der Hirsch* ('deer')
das Pony ('pony')	*der Rabe* ('raven')	*der Pinguin* ('penguin')
das Kamel ('camel')	*der Waschbär* ('racoon')	*der Fisch* ('fish')
	der Elefant ('elephant')	*der Hamster* ('hamster')
	der Eisbär ('bear')	*der Krebs* ('crab')
	der Löwe ('lion')	*der Papagei* ('parrot')
	der Falke ('hawk')	*der Fuchs* ('fox')

Of the 32 items, 16 were created for the SVO condition and 16 for the OVS condition, in which eight items represented the OVS early condition and eight items the OVS late condition. As reported in Section 5, OVS early sentences were disambiguated at the NP1 by accusative masculine case-marking and OVS late sentences were disambiguated at the NP2 on account of ambiguous neuter accusative nouns at the NP1 and unambiguous masculine nominative nouns at the NP2.

Thus, three different experimental conditions were presented (the use of strong and weak masculine nouns in the SVO word order had the only aim of balancing sentence conditions):

(1) SVO word order

a. nominative masculine nouns (NP1) and accusative strong masculine nouns (NP2) (4 items)
   Der Falke zieht den Hund
   [the hawk]NOM pulls [the dog]ACC
   'the hawk pulls the dog'

b. nominative masculine nouns (NP1) and accusative weak masculine nouns (NP2) (4 items)
   Der Wolf fotografiert den Elefanten
   [the wolf]NOM photographs [the elephant]ACC
   'the wolf photographs the elephant'

c. nominative neuter nouns (NP1) and accusative strong masculine nouns (NP2) (4 items)

   Das Küken grüßt den Krebs

   [the chick]$_{NOM}$ greets [the crab]$_{ACC}$

   'the chick greets the crab'

d. nominative neuter nouns (NP1) and accusative weak masculine nouns (NP2) (4 items)

   Das Nashorn malt den Pandabären

   [the rhinoceros]$_{NOM}$ paints [the panda]$_{ACC}$

   'the rhinoceros paints the panda'

Sentence conditions (1a) and (1b) were used to create their counterparts (2a) and (2b) for the OVS early sentence condition.

(2) OVS early

a. accusative weak masculine nouns (NP1) and nominative masculine nouns (NP2) (4 items)

   Den Falken pikst der Hund

   [the hawk]$_{ACC}$ pricks [the dog]$_{NOM}$

   'the dog pricks the haw'

b. accusative strong masculine nouns (NP1) and nominative masculine nouns (4 items)

   Den Wolf fesselt der Elefant

   [the wolf]$_{ACC}$ ties [the elephant]$_{NOM}$

   'the elephant ties the wolf'

To create the OVS late sentence condition, (1c) and (1d) were used (weak masculine nouns did not bear any overt inflections on the stem in the nominative case).

(3) OVS late

a. accusative neuter nouns (NP1) and nominative strong masculine nouns (NP2) (4 items)

   Das Küken impft der Krebs

   [the chick]$_{ACC}$ vaccinates [the crab]$_{NOM}$

'the crab vaccinates the chick'

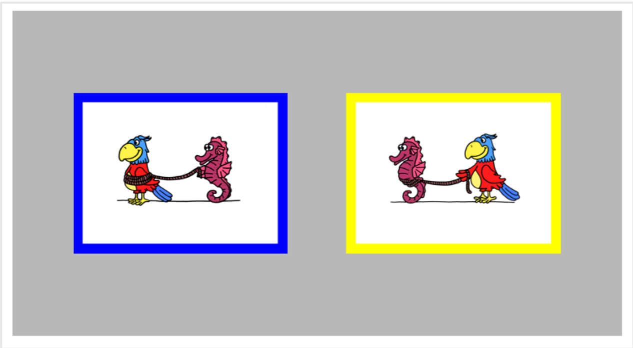
b. accusative neuter nouns (NP1) and nominative weak masculine nouns (NP2) (4 items)
Das Nashorn bespritzt der Pandabär
[the rhinoceros]ACC splashes [the panda]NOM
'the panda splashes the rhinoceros'

Altogether, eight action verbs were used in the experiment (i.e., the eight verbs reported to illustrate the experimental conditions in (1), (2) and (3)). Six items were created for the practice phase: the first two items contained a single word and the other four intransitive sentences. The experiment was divided into five different blocks with a short pause in between each. The first block was the practice phase. In each of the other four test blocks, four items were presented in the SVO and four in the OVS test condition. The same two animal characters were always mentioned together in an item, thus creating a pair. In the whole experiment they were each used twice with two different action verbs, i.e., once as agent of the action and once as patient. Each verb was mentioned once per block. On average, sentences were 2000 ms long: the mean length for NP1 was 750 ms, for the V 500 ms and for the NP2 750 ms. Two different versions of the experiment were created: the character having the agent role in one version had the patient role in the other one (see Appendix B for the complete versions). Across participants, the items were presented in a forward or backward order to avoid order effects. The order of trials was controlled for expected answers (half blue and half yellow), with no more than three consecutive same answers for the same color. All verbal stimuli were recorded by a female native speaker of German.

For each sentence, two pictures, each showing two characters involved in an action, were presented on the eye-tracking monitor screen (1920x1080 pixels in size). The two pictures of a pair only differed with respect to the roles of the two characters: while animal A was the agent performing an action on animal B in one picture, animal B performed the same action on animal A in the other picture. Overall, 64 single pictures were created. The picture presented on the left side of the screen always had a blue frame, while the picture on the right was shown with a yellow frame. Figure 4 represents the visual material for the item *Der Papagei fesselt das Seepferdchen* 'the parrot ties the seahorse'.

Figure 4

Example of visual stimuli array used in the OVS experiment



The pictures were colored with a free graphic tool and all the animal characters depicted were comparable in size. In half of the displays, the animal characters were oriented to the right side of the monitor, and in the other half to the left. The two pictures (each 500x354 pixels in size) were presented simultaneously side by side in the middle of the monitor on a white background, and the space between the two pictures was 150 pixels in size. Besides these areas, the background of the screen was gray in color (rgb color values: 184, 184, 184).

Procedure

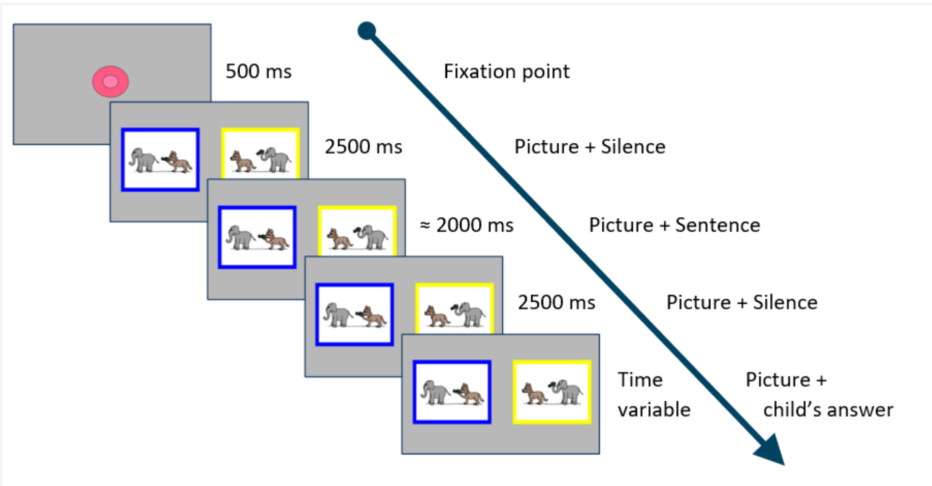
The eye-tracking system EyeLink Portable Duo (SR Research Ltd.) was used to measure each participant's eye-movements and pupil sizes with a tracking rate of 500 Hertz, thus data points were recorded every two milliseconds (ms). The right eye was recorded. Participants were seated in a quiet room at a distance of 50 to 55 cm from the EyeLink Portable Duo Eye Tracker, which was mounted below the screen of a display laptop. The eye-tracker operated in a head free-to-move remote mode and was connected to the host laptop (whose purpose was to integrate all the eye-tracking functionality) with a USB 3.0 cable. The experimenter sat next to the participant to control the quality of the tracking and start the stimuli presentation. Stimuli were presented using the SR Research Experiment Builder 2.1.140 Software (2017) in a dual-laptop set-up.

Prior to the beginning of the experiment, the drawings of the animals used in the experiments were shown to the children. After having heard the animal label, children needed to indicate the corresponding drawing, in order to show that they recognized the animals. During the adjustment of the eye-tracking system by the experimenter a music video for children was played. It was followed by a five-point calibration and five-point validation procedure displaying an animated frog on a black screen. If the validation was poor (less than one degree) the calibration was repeated. The cover story of the eye-tracking experiment was introduced by a girl, Anna, who wanted to play a game with the child.

The child was told that Anna showed pictures and said something about them. The child should decide which picture Anna was talking about by saying the color of the frame of this picture, namely blue or yellow. Figure 5 shows the structure of a single experimental trial. The trial started with a fixation point (500 ms) to attract the participant's attention followed by the verbal attention getter *Schau mal* ('Look at that'). After that the sentence condition was followed by the question *Welche Farbe hat der Rahmen* ('What color is the frame') in the first fourth practice items, in order to familiarize children with the task. Thereafter, the introductory question was not asked again (except to elicit a participant's answer after a few seconds' pause).

Figure 5

Experimental trial of the OVS experiment



After the disappearance of the fixation point, the two pictures were presented for around 7500 ms. The sentence began to play 2500 ms after the onset of the picture. The test sentences had an average duration of 2228 ms (ranging from 1501 – 2486 ms). After the end of the sentence, the pictures stayed on the screen for another 2500 ms. After the presentation of the sentence, children had no time restriction in which to label the color of the frame. Once children gave their responses, the experimenter started the next trial manually. The 38 trials (32 test trials and 6 practice trials) were presented in five blocks (i.e., one practice block and four test blocks) with short pauses in between, in which Anna appeared again and encouraged children to do well. The eye-tracking experiment was the first one performed by all children and lasted about 20 minutes.

5.1.2.4 Executive function tasks

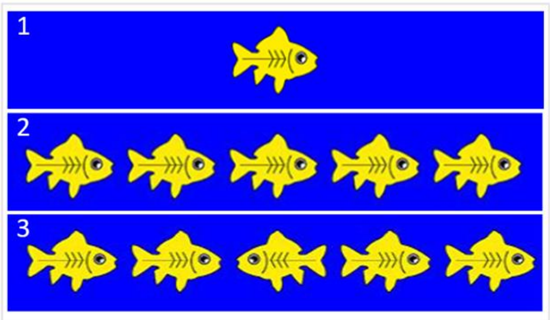
Two tasks were used to measure children's EFs: the Flanker task (Eriksen & Eriksen, 1974; Höhle et al., 2016; Rueda et al., 2004) and the task-switching paradigm (Jersild, 1927; Wiseheart et al., 2016).

<u>Flanker task</u>

The Flanker task was implemented on a laptop using OpenSesame (Mathôt et al., 2012). As illustrated in Figure 6, three different conditions were used: a neutral (1), a congruent (2) and an incongruent condition (3).

Figure 6

Visual representation of conditions in the Flanker task

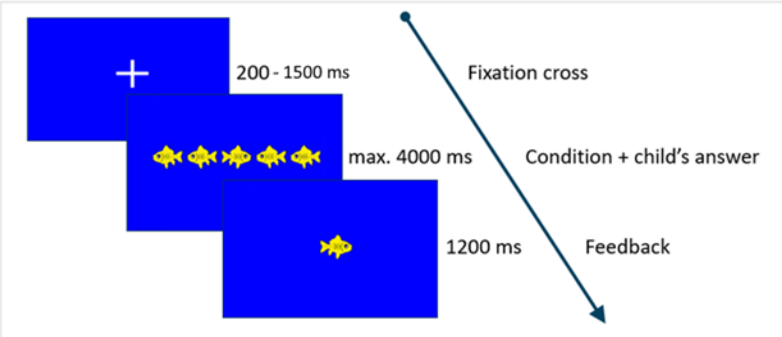


In the neutral condition, one fish appeared alone in the center of the screen. In the congruent condition, the middle fish was flanked by four other fish, two on each side, all looking in the same direction: either right or left. In the incongruent condition, the fish in the middle was also flanked by four other fish, but they looked in the opposite direction. The task consisted of indicating the direction in which the fish in the middle of the screen was looking.

Each experimental trial started with a fixation cross appearing in the middle of the screen with a random duration of between 200 and 1500 ms. The fixation cross was then replaced by a visual stimulus from one condition. The image remained on the screen until participants responded or for a maximum duration of 4000 ms. A feedback sound (about 1200 ms long) was played after each child's response (see Figure 7 for a presentation of a trial structure). Correct responses were followed by a winner sound and a swimming fish, while incorrect responses were followed by a bubbling sound and a static fish.

Figure 7

Trial structure in the Flanker task



All children were instructed to only pay attention to the fish in the center, even when more than one fish was shown, and to respond as fast as possible. Children were prompted to push a response button on the left or right side of the laptop, depending on the direction that the target fish was looking in. Thereby, the accuracy and RTs of children's responses were measured. The session started with six introduction items and two per conditions, which were used to explain the task to the children without time restriction. Thereafter, a practice phase started, including six practice items with a time limit of 4000 ms: four items were shown in the neutral condition, one in the congruent and one in the incongruent condition. If children were slower than 4000 ms or gave incorrect responses, the item was repeated. Before starting with the test phase, participants saw eight contours, in the form of white fish. This was used to visualize any progress during the experiment. Children were told that the game would be over when all the fish were painted. 48 test items with a time limit of 4000 ms were created and divided into eight blocks. Six items per block were used: two items per condition, once with the target fish pointing to the right and one pointing to the left. The experiment included no more than two consecutive trials of the same condition and no more than two consecutive trials with the fish looking in the same direction. All the fish were colored yellow and the same size. The task lasted about 10 minutes.

The Flanker task was used to measure the inhibition ability. When the direction of the flanking fish is inconsistent with the target, answering should require greater cognitive costs and thus longer RTs. As a measure for inhibition, the Flanker effect was used, namely the difference between the mean RT in the incongruent and congruent condition. The Flanker effect should be smaller for children with more advanced inhibition abilities.

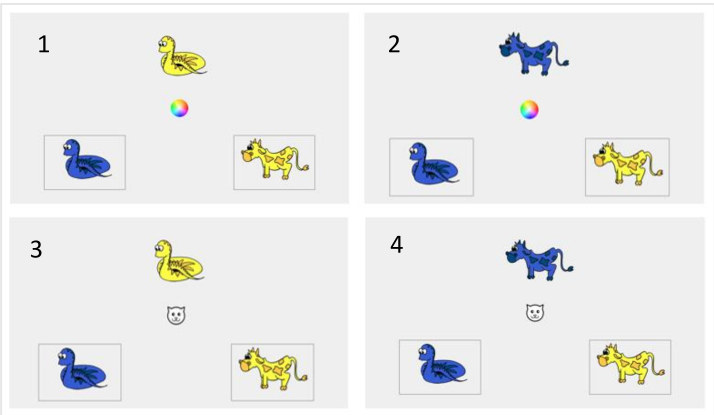
Task-switching paradigm

The task-switching paradigm was also implemented on a laptop using OpenSesame (Mathôt et al., 2012). The task-switching paradigm requires the ability to switch attention between two tasks. Two forms of costs are generally analyzed: global switch cost (GSC) and local switch cost (LSC). GSC is associated with brain regions that are believed to be involved in attention, goal-processing and planning. It reflects the activation of control mechanisms needed for maintaining two competing response sets and for a process of task decision in every trial. LSC is related to general response preparation linking task cues with the appropriate response mappings. Moreover, GSC reflects the need to resolve interference within a trial, while LSC is additionally driven by interference caused by the previous trial (Prior & MacWhinney, 2010; Wiseheart et al., 2016).

In the task, two response stimuli were shown at the bottom of the screen: a blue snake on the left side and a yellow cow on the right side. A task cue appeared in the middle of the screen: a colored wheel or an animal silhouette. The target stimulus appeared at the top of the screen above the task cue. Target stimuli were a yellow snake or a blue cow. As illustrated in Figure 8, four different combinations between cues and stimuli were possible.

Figure 8

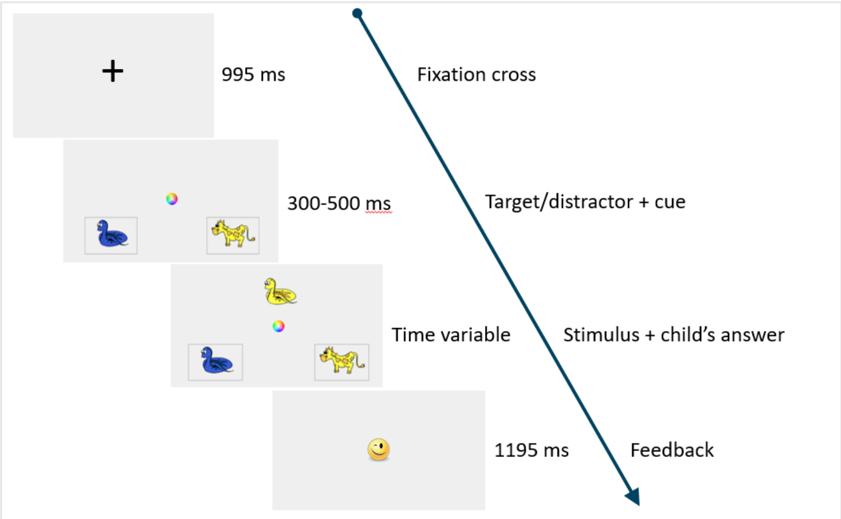
Illustration of the four combinations in the task-switching paradigm



In (1), the yellow snake above a color cue indicated the yellow cow as the correct response. In (2), the blue cow above a color cue indicated the blue snake as the correct response. In (3), the yellow snake above an animal cue indicated the blue snake as the correct response. In (4), the blue cow above an animal cue indicated the yellow cow as the correct response. Each trial started with a fixation cross presented for 995 ms in the center of the screen (Figure 9).

Figure 9

Trial structure in the task-switching paradigm



The fixation cross was replaced by the task cue, and at the same time the two response stimuli were presented at the bottom of the screen. To help children to perform the task, a verbal cue was added. This means that when the task cue graphically appeared on the screen, a preregistered voice integrated into the experiment announced the cue type, namely 'color' or 'animal'. The two response stimuli and the task cue remained on the screen for a random time of between 300-500 ms. Thereafter, the target stimulus appeared additionally at the top of the screen. The two response stimuli, namely the task cue and target stimulus, remained on the screen until the child responded. For negative responses, children heard an error sound and saw a sad emoticon. Positive feedback was used for correct responses: children heard a success sound and a smiling emoticon was shown. The length of the feedback sounds was 1195 ms.

First, children were introduced to the four characters of this task. During the introductory phase the experimenter explained to the children that they had to match the character appearing at the top of the screen to a character at the bottom of the screen, following the cue at the center of the screen. Four introduction trials were used to show participants the four different combinations between cues and stimuli. Two response buttons representing the two response stimuli were placed next to the laptop: the button showing the yellow cow was placed to the right of the laptop, and the button showing the blue snake was placed to the left, near the corresponding pictures on the screen. The children were prompted to press the button representing the target as fast as possible, during which time their accuracy and RT were measured. Before starting the test phase, they saw five uncolored winner emoticons. After the completion of each block of the test phase, children saw one of the five emoticons turn yellow. This was done to visualize their progress during the experiment. Participants were told that the game would be over when all the emoticons were yellow. At the end of the experiment, they heard an applause sound.

The test phase included five blocks. Two blocks were non-switch blocks and three were switch blocks. The two non-switch blocks (one with the color cue and one with the animal cue, counterbalanced across participants) included four practice trials followed by 10 non-switch trials. A trial was defined as a non-switch trial whenever the cue which had appeared in the previous trial replicated the cue of the current trial, namely from a trial with a color cue to a trial with a color cue, or from a trial with an animal cue to a trial with an animal cue. At the beginning of each block, one additional dummy trial was added which was not included in the analysis. Each stimulus, namely the yellow snake or the blue cow, was used in five trials. The three switch blocks included 20 trials: 10 non-switch trials and 10 switch trials. A trial was defined as a switch trial whenever the cue that had appeared in the previous trial did not replicate the cue of the actual trial, namely from a trial with a color cue to a trial with an animal cue, or from a trial with an animal cue to a trial with a color cue. Eight practice trials, four non-switch and four switch trials, were used before the first switch block was started. In the switch blocks each stimulus was used in 10 trials: five non-switch trials and five switch trials. The difference between mean RT for non-switch trials in the switch blocks and non-switch blocks was used to calculate the GSC. The difference between mean RT for non-switch and switch trials in the switch blocks was used to calculate LSC (Cepeda et al., 2001). Smaller GSC was previously associated with bilingual language experience, because it is assumed to account for processes more involved in executive control than LSC (Bialystok & Barac, 2012; Wiseheart et al., 2016). Therefore, the GSC was used in this study as a measure for switching ability.

The task included no more than three consecutive trials showing the same target stimulus (or the same cue in the case of switch blocks). The same expected answer was used for a maximum of two consecutive trials. Two versions of the task were created: in one version the task started with a non-switch block for color and in the other version it began with a non-switch block for animal. Each task lasted about 20 minutes.

5.1.2.5 Study sessions

Children were tested in two sessions, around two weeks apart. The first task of the first session was the sentence-picture verification task (≈ 20 minutes). Thereafter, children were tested using the visual memory subtest of the K-ABC II (≈ 10 minutes), the German TROG (≈ 15 minutes) and the Flanker task (≈ 10 minutes). During this first test session, bilinguals were additionally tested by means of the German PPVT (≈ 15 minutes). The Flanker task was the last one carried out for both groups. The first task of the second session for both groups was the task-switching paradigm (≈ 20 minutes). Thereafter, the monolingual children were tested by the digit span subtest of the K-ABC II (≈ 10 minutes), the CPM (≈ 10 minutes), and the German PPVT (≈ 15 minutes). The bilinguals were further also tested using the Italian PPVT (≈ 15 minutes) and the Italian TROG (≈ 15 minutes). Hence, the two testing sessions for monolingual children lasted around 45 minutes each, whereas for bilinguals they took around 60 minutes. All tests were conducted in German, except for the Italian PPVT and TROG.

5.1.3 Data analysis

For the statistical analysis, the R software version 4.2.1 (R Core Team, 2022) and its integrated development environment software application RStudio version 2022.7.2.576 (RStudio Team, 2022) were applied. For group comparisons of non-normal distributed data, the Mann-Whitney-U-Test was used, while t-tests were applied for data following a normal distribution. Linear mixed-effects models (LMM; Baayen et al., 2008) were calculated by using the lmer function of the lme4 package (D. Bates et al., 2015), and the *p*-values by using the R package lmerTest (Kuznetsova et al., 2017). LMMs allowed fixed and random effects to take the variation related to participants and items into account. Generalized LMMs were employed with binomial data using the glmer function of the lme4 package (D. Bates et al., 2015).

Eye-tracking data were filtered by the data analysis software SR Research Data Viewer 4.1.1 (2019) and uploaded into RStudio. Two spatial areas of interest (AoI) were defined. Each AoI was 500 x 354 pixels, corresponding to the size of the pictures without the frames. Proportions of looks were collapsed within participants and items for each time window. A time window was defined based on sentence constituents (NP1, V, NP2). The time windows for NP1 and NP2 were 750 ms long (i.e., the mean length of the NPs), and the time window for V was 500 ms (i.e., the mean length of verb). This was carried out for all sentences, meaning that an average time window was used for the analysis. After sentence offset, three different time windows (Sil1, Sil2, Sil3) were considered, which all included 750 ms of silence.

Correct responses in the sentence comprehension task were coded as 1 and incorrect responses as 0. The independent variables for the offline and online methods were sentence condition (SVO/OVS early/OVS late) and language group (monolinguals/bilinguals). The dependent variables were the accuracy of correct responses, the proportion of looks to the target picture across the different time windows, or the pupil dilation, depending on the experiment.

5.2 Results

5.2.1 Language exposure and socio-economic profile

The cumulative language input and output were calculated. The estimated number of hours that parents spent with their child in a week and to what percentage they spoke German with the child (following Unsworth, 2013) were multiplied to calculate the number of hours that parents spoke German with the child in a week. This was repeated for all sources of input asked in the questionnaire (siblings, babysitter, kindergarten, reading out, TV/radio/PC and sport). Summing up these hours of German input and dividing by overall hours of input, it calculates the child's individual percentage of German input as a weighted average. Percentage scores of German output by each child were calculated by the same procedure used for the input. For five children, the number of hours spent with their parents was not reported. Parents only wrote the type of spoken language and its relative use in percentage. So, assuming that these children also spent time with their parents, missing values were replaced with scale values that accounted for the mean and standard deviation (SD) at the group level.

A score for the parents' education was also calculated. Since there was a positive correlation between mothers' and fathers' education (*rho* = .513, *p* = .003), a mean value of these was used as the parent's education variable. Table 4 describes the mean, SD, median, minimum (Min) and maximum (Max) values (i.e., range) for the average proportion of cumulative input and output in German and parents' education of bilingual children.

Table 4

Descriptive statistics of covariates for the bilingual group

	Mean	SD	Median	Min	Max
Proportion cumulative input in German (%)	59.9	13.8	61.7	23.7	85.3
Proportion cumulative output in German (%)	48.9	18.7	44.0	11.1	84.3
Parents' education	4	0.7	4	2	5

At the group level, children had a relatively balanced exposure to German and Italian and spoke both languages. Nevertheless, a considerable individual variation is visible observing the Min and Max values, namely the range: while one child heard German on average 23.7% of the time, another child heard German 85.3% of the time. For language output, the range was even larger: 11.1%-84.3%. On a scale from 1 to 5, bilinguals' parents had on average a level of education of 4. The parents' education was also calculated for the monolingual children (mean: 3.7; SD: 0.6; median: 4; range: 2-4.5) and used as variable for further analyses, as for the bilingual group. These data were calculated from the language background questionnaire that parents of the monolingual children had completed.

Finally, the number of older siblings was calculated for the bilingual (mean: 0.75; SD: 0.54; median: 1; range: 0-2) and the monolingual group (mean: 0.37; SD: 0.81; median: 0; range: 0-3). The sample of 31 bilingual children consisted of eight only children, 12 children with younger siblings (each child with only one younger sibling) 9 children with older siblings (six children with one older sibling and three with two older siblings) and missing data for two children. All older siblings were school-aged. The sample of 31 monolingual children included: eight only children, five children with younger siblings (three children with only one younger sibling and two children with other two younger siblings), seven children with older siblings (three children with one older sibling, one child with two older siblings and one younger sibling, and three with one older and one younger sibling) and missing data for eleven children (i.e., not reported in the questionnaire). All older siblings were school-aged.

Additionally, two variables from the language background questionnaire were also compared between the two groups: the parents' education and the number of older school-aged siblings. The Mann-Whitney-U-test showed a higher level of parents' education in the bilingual families in comparison to monolingual families (W = .671, p = .004) and no difference in the number of older siblings between the two groups (W = 284.5, p = .892). The number of older siblings, as well as language experience (i.e., cumulative input and output) and the parents' education, were entered as the covariate in a subsequent analysis (see Section 5.2.6 for these results).

5.2.2 Standard assessments

Monolinguals and bilinguals were compared on their performance in the standardized tests. To test normality, the Shapiro-Wilk test was used. This test showed that the data from the German PPVT (W = .975, p = .240), visual memory (W = .969, p = .119) and CPM (W = .968, p = .107) followed a normal distribution, while the data from the German TROG (W = .939, p = .004) and digit span (W = .938, p = .004) were not normally distributed. Unpaired t-tests and Mann-Whitney-U-Tests were respectively used. Table 5 shows the descriptive data for all standardized tests in both language groups. Mean, SD, median, Min and Max values are given.

Table 5

Performance of children in the standardized tests by group comparisons

		Mean	**SD**	**Median**	**Min**	**Max**
PPVT (German) (max. score 228)						
	Monolingual children	128	17.2	128	87	155
	Bilingual children	123	22.1	123	83	174
TROG (German) (max. score 84)						
	Monolingual children	66.2	11.1	70	37	79
	Bilingual children	62.4	10.9	62	30	80
Digit span (K-ABC II) (max. score 21)						
	Monolingual children	8.4	2	8	5	12
	Bilingual children	8.5	1.8	8	5	11
Visual memory (K-ABC II) (max. score 31)						
	Monolingual children	14.2	2.8	15	7	19
	Bilingual children	13.8	3.3	13	9	23
CPM (max. score 36)						
	Monolingual children	18.6	4.7	18	9	30
	Bilingual children	18.9	3.5	19	12	29

Unpaired t-tests showed no significant differences between the performance of the monolingual and bilingual groups in the German PPVT ($t(57) = 1.003$, $p = .320$), visual memory ($t(59) = .495$, $p = .622$) and CPM ($t(56) = -.274$, $p = .785$) tests. The Mann-Whitney-U-test showed no significant differences between the performance of both groups in the German TROG ($W = 360$, $p = .088$) and digit span ($W = 495$, $p = .841$) tests.

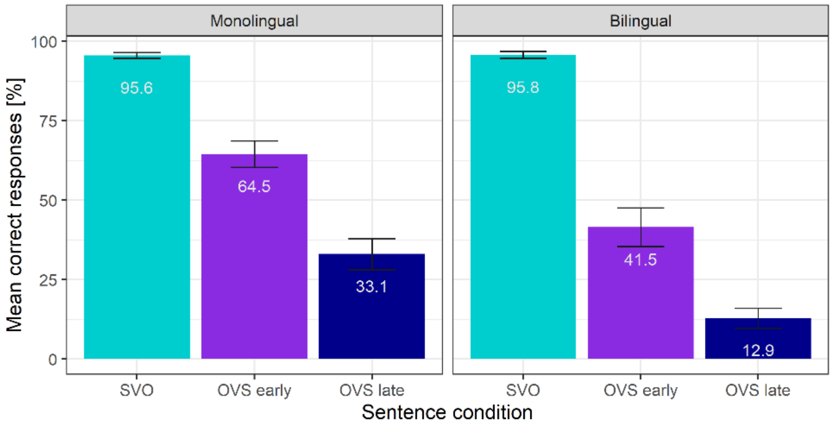
5.2.3 Sentence comprehension and processing

5.2.3.1 Accuracy

The mean percentage of correct responses of children in the SVO, OVS early and OVS late sentence conditions are presented in the next Figure.

Figure 10

Mean percentage of correct responses in sentence comprehension task



Note. Error bars present ±1 standard error (SE) of between-subject variance.

Children performed close to ceiling in the SVO sentence condition. One-sample t-tests were run to compare children's performance in the OVS conditions against chance level performance (i.e., 50% od correct responses). In the OVS early sentence condition, bilingual children performed at chance level ($t(30) = -1.397$, $p = .173$) and monolingual children performed above chance level ($t(30) = 3.470$, $p = .002$). In the OVS late condition, both language groups performed below chance level (bilingual group: $t(30) = -11.786$, $p < .001$, and monolingual group: $t(30) = -3.454$, $p < .002$).

As mentioned in Section 5.1.2.3, OVS early condition contained weak and strong accusative masculine nouns (remember that the OVS late condition did not include weak nouns, since all masculine nouns were in the nominative and not in the accusative case-marking). Table 6 reports the related mean accuracy score for the monolingual and bilingual groups, respectively.

Table 6

Mean percentage of correct responses for weak and strong masculine nouns in OVS early sentences

	OVS early condition	
	strong nouns	**weak nouns**
Monolingual children	66.9%	62.1%
Bilingual children	45.2%	37.9%

Paired *t*-tests revealed no differences between strong and weak nouns in the OVS early condition (all $t < 1.438$, all $p > 0.161$), neither in the monolingual nor in the bilingual group.

The mean number of correct responses by the 31 bilingual children were compared to the data from the 31 monolingual children using a LMM with treatment contrasts for the factor condition (SVO/OVS early/OVS late), as well as for the factor language group (i.e., bilinguals/monolinguals, henceforth Bili/Mono), in which subsets of the data, in this case monolinguals and OVS early, were coded to be in the intercept. The dependent variable was the number of correct responses. The model included two random intercepts: participants and items (i.e., 64 pictures created for the experiment), and a random slope: individual adjustment of condition effect for each participant (following Barr et al., 2013) to specify a maximal random effects structure for confirmatory hypothesis testing. Table 7 illustrates parameters of the LMM: estimates, SE, *z*-values and *p*-values are given for fixed factors (see Appendix C.1 for complete model output). Note that under the fixed effect column I placed the subset of data in brackets, to which the simple effect refers. This will be the case for all further analyses.

Table 7

Parameters of LMM for correct responses by monolinguals and OVS early in the intercept

Fixed effect	**Estimate**	**SE**	***z*-value**	***p*-value**
Intercept (Mono, OVS early)	0.821	0.295	2.785	=.005**
OVS early (Bili)	-1.262	0.405	-3.120	=.002**
OVS late (Mono)	-1.705	0.282	-6.054	<.001***
SVO (Mono)	2.601	0.374	6.949	<.001***
Bili x OVS late	-0.396	0.407	-0.975	=.330
Bili x SVO	1.440	0.522	2.757	=.006**

** $p < .01$, *** $p < .001$

Estimates indicate the difference scores of the logit-transformed proportion of correct responses (zero = 1:1 = 50%). The number of correct responses for OVS early in monolinguals was above chance (*p* = .005) and bilinguals gave significantly fewer correct responses than monolinguals in this condition (*p* = .002). Within the monolingual group and with respect to the OVS early condition, fewer correct responses were given in the OVS late condition (*p* < .001) and more correct responses were given in the SVO condition (*p* < .001). The decrease of correct responses to OVS late vs. OVS early for bilinguals (28.6%) was comparable (*p* = .33) to that of monolinguals (31.4%), while the increase of correct responses to SVO vs. OVS early was significantly larger (*p* = .006) for bilinguals (54.3%), compared to monolinguals (31.1%).

A post hoc comparison was made to evaluate whether the number of correct responses in the bilingual group was different between conditions and if there were any differences between the groups' performance in the OVS late sentence condition. Therefore, a further LMM with different contrast coding (bilinguals and OVS late coded to be in the intercept) was run. Further model specifications were the same as in the previous model. Table 8 reports the results of the model (complete output in Appendix C.2).

Table 8

Parameters of LMM for correct responses with bilinguals and OVS late in the intercept

Fixed effect	**Estimate**	**SE**	***z*-value**	***p*-value**
Intercept (Bili, OVS late)	-2.540	0.334	-7.609	<.001***
OVS late (Mono)	1.656	0.426	3.892	<.001***
OVS early (Bili)	2.100	0.331	6.343	<.001***
SVO (Bili)	6.140	0.473	12.979	<.001***
Mono x OVS early	-0.395	0.406	-0.972	=.331
Mono x SVO	-1.835	0.625	-2.934	=.003**

** *p* < .01, *** *p* < .001

The number of correct responses for OVS late in bilinguals was below chance level (*p* < .001), and monolinguals gave significantly more correct responses than bilinguals in this condition (*p* < .001). Within the bilingual group and with respect to the OVS late condition, more correct responses were given for the OVS early condition (*p* < .001) and for the SVO condition (*p* < .001). Interaction effects reflected the previous model: the decrease of correct responses to OVS early vs. OVS late for monolingual children is comparable to that of bilingual children (*p* = .331), while the increase of correct responses to SVO vs. OVS late is significantly smaller (*p* = .003) for monolinguals compared to bilinguals.

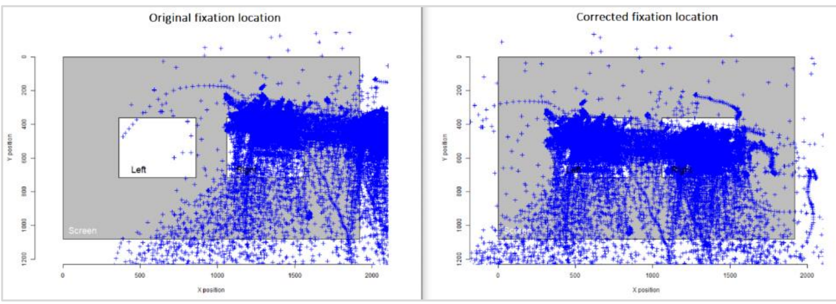
Taken together, these results showed an effect of word order: SVO sentences were well performed in both language groups, OVS early sentences were performed above chance in the monolingual group and at chance level in the bilingual group, and OVS late sentence were performed below chance level in both groups. SVO sentences were performed better than OVS early sentences and OVS early were performed better than OVS late sentences in both monolinguals and bilinguals. Monolinguals gave significantly more correct responses than bilinguals in the comprehension of both types of OVS sentences, but not of SVO sentences.

5.2.3.2 Pre-processing of eye-gaze data

Samples with blinks or gaze positions outside the screen were removed. The remaining data for the whole group contained an average of 87.8% (range: 35.8%-95%) looks to the AoIs (47.6% to the target and 40.2% to the distractor). The quality of the eye-tracking data, especially the spatial precision, depends on the calibration at the very beginning. This in turn is heavily influenced by the sitting position (head location, distance of the eyes to the screen) and might change in the course of the experiment, particularly with young children – even for well-calibrated participants (Cohen, 2013). Therefore, an offline calibration check inspired by Frank et al. (2012) was performed. Using the central fixation point at the beginning of each trial, the median offset for each participant's X and Y positions during that time was measured (averaged for the whole experiment), and this value was used to calculate a corrected gaze position. The original and corrected gaze positions of each participant were then plotted (for an example, see Figure 11).

Figure 11

Example of original and corrected recorded eye fixation location of one child



Note. The x-axes represent the X coordinates of eye gaze positions on the screen and the y-axes the Y coordinates recorded during the eye-tracking experiment. Original fixation location is represented on the left image and corrected on the right.

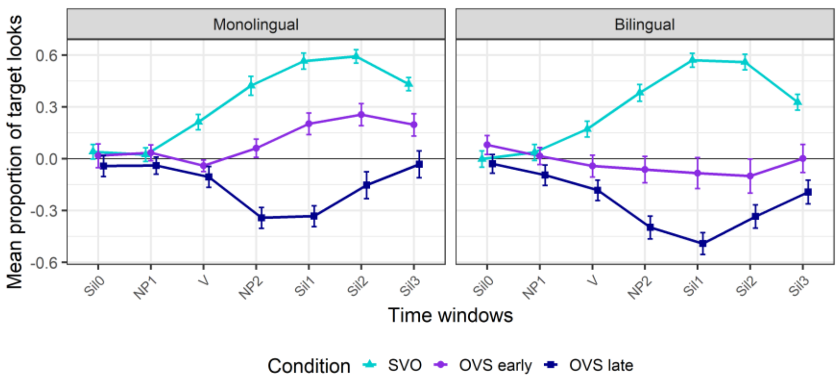
As Figure 11 above illustrates, a change of the eye fixation location is easy to detect visually in the plots. A criterion of substantial improvement was used to decide whether the original or the corrected gaze positions were to be considered for the analysis: if the number of data points falling into one of the two AoIs (target or distractor picture excluded the frame) increased by 10% or more in the corrected compared to the original gaze positions, then the corrected gaze positions were included in the analysis (note that this procedure is blind to accuracy and condition). This was only the case for one of the children (reported in Figure 11). The original fixation positions of this child into the AoIs represented 40.6% of the total and the corrected fixation positions 90.4%. In 15 other children, the increase in gazes to the AoI was less than 1.7% and in three other children, the correction led to an increase of eye fixation of 2.6%, 4.3% and 8.7%. For all of these children, the original gaze positions were used according to the set criterion.

5.2.3.3 Eye-gaze data

Mean proportion of looks to the target picture was calculated by dividing the difference between target looks minus distractor looks by the sum of target looks and distractor looks. The eye-gaze data (proportion of looks to the target) were first averaged for each time window (levels: NP1, V, NP2, Sil1, Sil2, Sil3) in every trial, across all trials of each condition (SVO, OVS early, OVS late) and each participant and subsequently across participants within each group (monolinguals and bilinguals). Sil0 indicated the time of 750 ms of silence until the onset of NP1, time window NP1 indicated the mean length between NP1 onset and NP1 offset, time window V the mean length between V onset and V offset and so on. Figure 12 plots the mean proportion of looks to the target (y-axis) for SVO, OVS early and OVS late sentences during the six time windows (x-axis).

Figure 12

Proportion of looks to the target for SVO, OVS early and OVS late conditions separated by group



Note. The left panel shows the proportion of target looks for each sentence condition in the monolingual group and the right panel in the bilingual group. The gray horizontal line at 0 indicates chance level. Error bars represent ±1 SE of between-subject variance.

LMMs with participants and items as random effects were fitted. Random slopes were not specified since their inclusion led the model to non-convergence. Fixed effects included language group and condition, as well as their interaction. The looks to the target picture for each time window were modelled separately. To compare the proportion of looks to the target across different conditions, a treatment contrast was applied. Hence, the monolingual group and the OVS early sentence condition were first set in the intercept. The intercept represented the chance level of monolingual children in the OVS early sentence condition. Table 9 provides the parameters for the LMMs (see Appendix C.3-C.8 for complete model outputs). Estimates, SE, degree of freedom (df), *t*-values and *p*-values are given.

Table 9

Parameters of LMMs for eye-gaze data with monolinguals and OVS early in the intercept

Time window	Fixed effect	Estimate	SE	df	*t*-value	*p*-value
NP1	Intercept (Mono, OVS early)	0.037	0.058	1832	0.632	=.527
	OVS early (Bili)	-0.031	0.083	1832	-0.375	=.707
	OVS late (Mono)	-0.076	0.082	1832	-0.928	=.354
	SVO (Mono)	-0.016	0.072	1832	-0.218	=.827
	Bili x OVS late	-0.044	0.116	1832	-0.374	=.708
	Bili x SVO	0.049	0.101	1832	0.486	=.627
V	Intercept (Mono, OVS early)	-0.044	0.064	187.4	-0.689	=.492
	OVS early (Bili)	-0.009	0.086	1761	-0.110	=.912
	OVS late (Mono)	-0.065	0.090	181.8	-0.723	=.471
	SVO (Mono)	0.262	0.078	186.4	3.348	<.001***
	Bili x OVS late	-0.066	0.120	1762	-0.545	=.586
	Bili x SVO	-0.021	0.105	1763	-0.202	=.840
NP2	Intercept (Mono, OVS early)	0.060	0.061	157.12	0.983	=.327
	OVS early (Bili)	-0.139	0.079	689.16	-1.761	=.079‡
	OVS late (Mono)	-0.409	0.085	167.82	-4.806	<.001***
	SVO (Mono)	0.372	0.074	170.57	5.016	<.001***
	Bili x OVS late	0.095	0.011	1730	0.861	=.389
	Bili x SVO	0.100	0.096	1730	1.042	=.298
Sil1	Intercept (Mono, OVS early)	0.207	0.061	167.35	3.410	<.001***

Time window	Fixed effect	Estimate	SE	df	*t*-value	*p*-value
	OVS early (Bili)	-0.290	0.077	472.49	-3.746	<.001***
	OVS late (Mono)	-0.544	0.081	161.32	-7.488	<.001***
	SVO (Mono)	0.366	0.071	164.43	5.184	<.001***
	Bili x OVS late	0.143	0.103	1741	1.395	=.163
	Bili x SVO	0.299	0.089	1741	3.357	<.001***
Sil2	Intercept (Mono, OVS early)	0.255	0.061	181.41	4.186	<.001***
	OVS early (Bili)	-0.366	0.083	364.63	-4.412	<.001***
	OVS late (Mono)	-0.408	0.078	197.86	-5.238	<.001***
	SVO (Mono)	0.337	0.068	198.05	4.987	<.001***
	Bili x OVS late	0.185	0.105	1723	1.754	=.080‡
	Bili x SVO	0.337	0.091	1725	3.693	<.001***
Sil3	Intercept (Mono, OVS early)	0.195	0.060	163.12	3.234	<.001***
	OVS early (Bili)	-0.209	0.084	564.85	-2.474	=.014*
	OVS late (Mono)	-0.218	0.083	191.40	-2.623	=.009**
	SVO (Mono)	0.241	0.072	193.89	3.343	<.001***
	Bili x OVS late	0.051	0.115	1660	0.443	=.658
	Bili x SVO	0.100	0.100	1664	0.995	=.320

‡ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$

During time window NP1, the proportion of looks to the target was similar between conditions and language groups. In the time windows V, NP2, Sil1, Sil2 and Sil3, however, the proportion of looks to the target was smaller in OVS early as compared to SVO (all $p < .001$) for the monolingual group. In the time windows NP2, Sil1, Sil2 and Sil3, the proportion of looks to the target was smaller in OVS late as compared to OVS early ($p < .001$ for NP2, Sil1 and Sil2; $p = .009$ for Sil3) again for the monolingual group. In Sil1, Sil2 and Sil3 monolinguals showed larger proportion of looks to target as compared to bilinguals ($p < .001$ for Sil1 and Sil2 and $p = .014$ for Sil3) in OVS early. Significant interaction effects in Sil1 and Sil2 (Bili x SVO) indicated that the contrast between OVS early and SVO sentences was smaller for the monolingual than for the bilingual children. From Sil1, monolingual children started looking to the target above chance level until Sil3 in OVS early sentences.

Similarly to the analysis with the number of correct responses, a post hoc comparison was made to evaluate whether the proportion of target looking in the bilingual group was different between conditions in each time window and also whether the looking behavior of groups was different during the processing of OVS late sentences. Therefore, other LMMs with different contrast coding were run (bilinguals and OVS late coded to be in the intercept). Further model specifications were identical to those in the previous models. Table 10 reports these results (complete outputs in Appendix C.9-C.14).

Table 10

Parameters of LMMs for eye-gate data with bilinguals and OVS late in the intercept

Time window	**Fixed effect**	**Estimate**	**SE**	**df**	***t*-value**	***p*-value**
NP1	Intercept (Bili, OVS late)	-0.113	0.059	1832	-1.929	=.054‡
	OVS late (Mono)	0.075	0.082	1832	0.910	=.363
	OVS early (Bili)	0.119	0.083	1832	1.438	=.151
	SVO (Bili)	0.153	0.072	1832	2.122	=.034*
	Mono x OVS early	-0.044	0.116	1832	-0.374	=.709
	Mono x SVO	0.093	0.101	1832	-0.921	=.357
V	Intercept (Bili, OVS late)	-0.184	0.064	192.72	-2.862	=.005**
	OVS late (Mono)	0.075	0.085	1762	0.887	=.375
	OVS early (Bili)	0.131	0.091	193.43	1.433	=.153
	SVO (Bili)	0.372	0.079	193.38	4.713	<.001***
	Mono x OVS early	-0.066	0.120	1762	-0.545	=.586
	Mono x SVO	-0.044	0.104	1763	-0.428	=.669
NP2	Intercept (Bili, OVS late)	-0.393	0.061	165.3	-6.398	<.001***
	OVS late (Mono)	-0.044	0.079	680.2	0.561	=.575
	OVS early (Bili)	0.314	0.087	183	3.615	<.001***
	SVO (Bili)	0.785	0.075	177.7	10.521	<.001***
	Mono x OVS early	0.095	0.110	1730	0.861	=.389
	Mono x SVO	-0.005	0.095	1724	-0.048	=.961
Sil1	Intercept (Bili, OVS late)	-0.484	0.061	169.92	-7.975	<.001***
	OVS late (Mono)	0.147	0.077	460.38	1.915	=.056‡
	OVS early (Bili)	0.410	0.082	168.85	4.911	<.001***
	SVO (Bili)	1.065	0.071	168.36	15.07	<.001***

Time window	Fixed effect	Estimate	SE	df	*t*-value	*p*-value
	Mono x OVS early	0.143	0.103	1741	1.395	=.163
	Mono x SVO	-0.155	0.089	1743	-1.761	=.079‡
Sil2	Intercept (Bili, OVS late)	-0.334	0.061	180.93	-5.514	<.001***
	OVS late (Mono)	0.181	0.083	357.84	2.197	=.029*
	OVS early (Bili)	0.223	0.078	201.29	2.862	=.005**
	SVO (Bili)	0.897	0.068	199.06	13.32	<.001***
	Mono x OVS early	0.185	0.105	1723	1.754	=.080‡
	Mono x SVO	-0.152	0.091	1721	-1.677	=.094‡
Sil3	Intercept (Bili, OVS late)	-0.180	0.062	179.69	-2.913	<.001***
	OVS late (Mono)	0.158	0.085	569.87	1.861	=.063‡
	OVS early (Bili)	0.167	0.085	209.49	1.965	=.051‡
	SVO (Bili)	0.507	0.073	205.89	6.936	<.001***
	Mono x OVS early	0.051	0.115	1660	0.443	=.658
	Mono x SVO	-0.048	0.100	1653	-0.483	=.629

‡ $p < .1$, * $p <.05$, ** $p <.01$, *** $p < .001$

From the time window NP1 until Sil3, the proportion of looks to the target was larger in SVO as compared to OVS late for the bilingual group ($p = .034$ for NP1 and $p < .001$ for all other time windows). In the time windows NP2, Sil1, Sil2 and Sil3, the proportion of looks to the target were smaller in OVS late as compared to OVS early ($p < .001$ for NP2 and Sil1; $p = .005$ for Sil2 and $p = .051$ for Sil3), again for the bilingual group. In Sil2, monolinguals showed larger proportion of looks to the target as compared to bilinguals ($p = .029$) in OVS late. From V, bilingual children started looking to the target below chance level until Sil3 in OVS late sentences.

Now, to evaluate the looking behavior of monolingual children during OVS late sentences and the looking behavior of bilingual children during OVS early sentences with respect to the chance level, two further post hoc analyses were performed (see Appendix C.15 and Appendix C.16). The proportion of looks to the target in OVS late sentences was below chance from NP2 up to Sil2 in the monolingual group (all $p < .001$). During the time window Sil3, the proportion of looks to the target by monolinguals was at chance level. In the bilingual group, the proportion of looks to the target was smaller in OVS early, compared to SVO from time window V up to Sil3 (all $p < .01$). The proportion of looks to the target in OVS early sentences was at chance level across all time windows in the bilingual group.

In summary, both groups looked more to the target during SVO than OVS early sentences and more during OVS early than OVS late sentences. Specifically, the proportion of looks to the target was larger during SVO than OVS early from the V until Sil3 for both groups. The proportion of looks to the target was larger during OVS early than OVS late from the NP2 until Sil3 for both groups. Differences between the groups were visible after sentence offset in both OVS sentences: the proportion of looks to the target by the monolingual group was larger as compared to the bilingual group from Sil1 until Sil3. In OVS late sentences, the proportion of looks to the target by the monolingual group was only larger as compared to the bilingual group during Sil2. From Sil1, the monolinguals' proportion of looks to the target was above chance in OVS early sentences and only at chance in Sil3 in OVS late sentences. Bilinguals' proportion of looks to the target was at chance in all the time windows while processing OVS early sentences and always below chance in OVS late sentences.

Additionally, in order to investigate whether participants' gaze data predict their responses or show inconsistency between offline and online results (as e.g., in Brandt-Kobele & Höhle, 2010; Höhle et al., 2016; Sekerina et al., 2004), LMMs including the factor responses were run (correct and incorrect). Before starting with the models, individual data were explored. Table 11 reports the number of participants given a specific number of correct responses in the SVO condition.

Table 11

Individual raw data for correct responses in the SVO condition divided by group

	SVO condition	
Responses	**Monolinguals**	**Bilinguals**
16/16 corr. resp (100%)	16	20
15/16 corr. resp (93.75%)	9	5
14/16 corr. resp (87.5%)	5	4
13/16 corr. resp (81.25%)	1	1
12/16 corr. resp (75%)	-	-
11/16 corr. resp (68.75%)	-	1

Note. The total number of correct responses in the SVO condition in monolinguals was 474 items out of 496 (95.6%) and 475 out of 496 items (95.8%) in bilinguals.

Table 12 indicates the number of participants given a specific number of correct responses in the OVS early and OVS late condition.

Table 12

Individual raw data for correct responses in the OVS early and OVS late condition divided by group

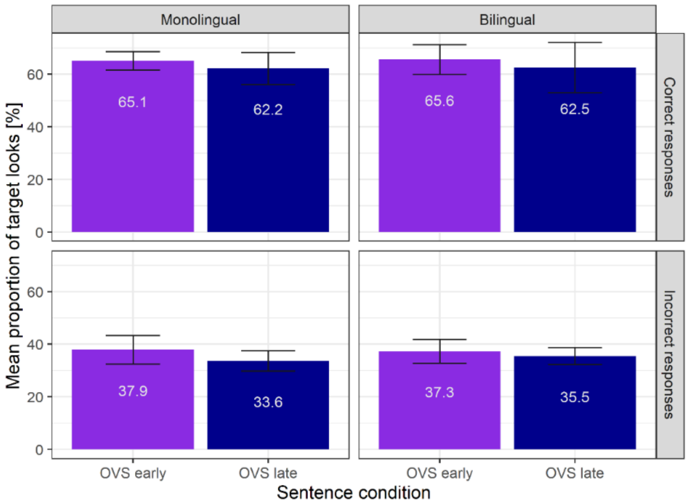
	OVS early condition		**OVS late condition**	
Responses	**Monolinguals**	**Bilinguals**	**Monolinguals**	**Bilinguals**
8/8 corr. resp (100%)	4	4	1	-
7/8 corr. resp (87.5%)	4	1	2	-
6/8 corr. resp (75%)	5	3	2	-
5/8 corr. resp (62.5%)	8	3	1	-
4/8 corr. resp (50%)	4	1	-	2
3/8 corr. resp (37.5%)	2	3	8	2
2/8 corr. resp (25%)	4	6	5	4
1/8 corr. resp (12.5%)	-	6	9	8
0/8 corr. resp (0%)	-	4	3	15

Note. The total number of correct responses in the OVS early condition in all 31 monolinguals is 160 out of 248 (64.5%) and 103 out of 248 (41.5%) in 27 bilinguals, since four bilinguals gave no correct answers. In the OVS late condition, the number of correct responses by 28 monolinguals is 82 out of 248 (33.1%) and in 16 bilinguals 32 out of 248 (12.9%).

The performance of all children, except one bilingual child, was at least of 81.25% in the SVO condition. The data for correct responses in the OVS early condition demonstrate that approx. 19% of monolingual children and approx. 81% of bilingual children were below 50% of accuracy. In the OVS late condition, the data prove that approx. 19% of monolingual children gave more than 50% of correct responses, while every bilingual child gave exactly or less than 50% of correct responses. Thus, the SVO sentence condition was removed from the next model, on account of the at near ceiling performance and low variance. Whether participants' eye-gaze data were related to their responses was subsequently investigated: the factor accuracy (i.e., correct/incorrect responses) for the OVS early and the OVS late conditions was added as a predictor. Contrary to previous models, the mean proportion of target looking was used here as the dependent variable across all time windows (Sil0 was excluded because it included 750 ms before sentence onset). This was done because the aim of this analysis was to show potential differences between offline (responses) and online (looking patterns) data over the entire trial. Repeated contrast coding was used for both fixed factors. In the contrast OVS late-OVS early, negative estimates indicated more looks to the target for OVS early, while in the contrast Bili-Mono they indicated more looks to the target for monolinguals. Trials with correct responses were coded to be in the intercept. Thus, fixed effects of the next LMM were language group, condition and accuracy. Random effects were three random components: participants and items as random intercepts and individual adjustment of condition per participant as random slope. Results of this LMM are displayed in the next bar plot (Figure 13).

Figure 13

Proportion of target looking for OVS early and OVS late, broken down by group and responses



Note. The mean proportion of looking to the target is shown on the y-axis and the OVS sentence conditions are reported on the x-axis. Error bars represent ±1 SE of between-subject variance.

Next Table reports the results for this LMM model (see Appendix C.17 for complete output).

Table 13

Parameters for the LMMs for eye-gaze data with correct responses in the intercept

Fixed effect	**Estimate**	**SE**	**df**	***t*-value**	***p*-value**
Intercept (Corr. resp, grand mean)	0.645	0.014	143.4	46.636	<.001***
OVS late-OVS early	-0.044	0.029	154.2	-1.528	=.129
Bili-Mono	0.0004	0.025	326.3	0.015	=.988
Incorr. resp	-0.311	0.014	1353	-21.456	<.001***
OVS late-OVS early x Bili-Mono	-0.002	0.052	289.6	-0.033	=.974
OVS late-OVS early x Incorr. resp	0.013	0.029	1787	0.434	=.664
Bili-Mono x Incorr. resp	-0.007	0.029	1374	-0.251	=.802
OVS late-OVS early x Bili-Mono x Incorr. resp	0.037	0.059	1812	0.631	=.528

*** p < .001

When incorrect responses were given, children looked less to the target in comparison to their correct responses ($p < .001$) in both OVS conditions. This means that inconsistency between offline and online data did not appear.

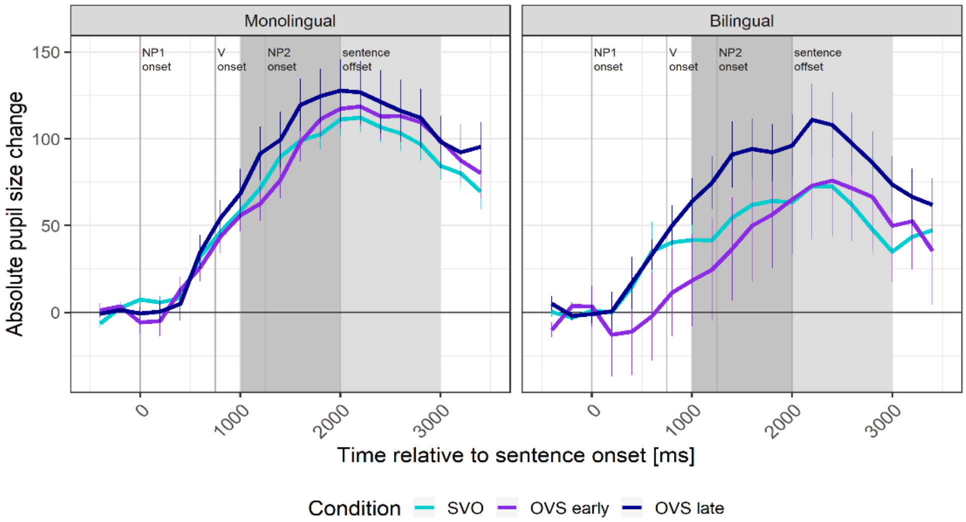
5.2.3.4 Pupillometry

Pupil size was also measured by the EyeLink eye tracker system and reported as pupil area recorded in scaled camera image pixels. Typical pupil area is between 100 and 10000 arbitrary units[6]. The eye-tracking system provides pupil data in arbitrary units, since pupil data are not calibrated and the units of pupil measurement vary according to individual calibration for different participants. In fact, pupil data are affected by up to 10% by pupil position, due to the optical distortion of the cornea that accompanies rotations of the eye and camera-related factors. Pupil dilation was expressed as proportional change relative to a baseline window. This baseline pupil dilation was taken from the interval 500 msec prior to the sentence onset for each trial and subtracted from the subsequent data points. Trials displaying more than 25% missing data were removed, which affects 13.7% of the data. Eight out of 62 children concluded with less than 25% missing pupil data in all trials. On average, five trials per participant had to be removed (range: 1-18 trials). Missing points were not interpolated. Considering that the pupil reaches its maximum deflection between one and two seconds after the evoking event (Just & Carpenter, 1993) and that children may need more time to reach maximum pupil size, two time windows were analyzed. Time window 1 was set to between 1000 ms and 2000 ms because the first disambiguating cue was sentence initial on the NP1, and time window 2 was set to between 2000 ms and 3000 ms because the earliest onset of NP2 was at 1073 ms from sentence onset and the latest onset of NP2 was at 1702 ms from sentence onset. Thus, contrary to eye-gaze data in which the mean proportion of looks to the target was calculated between the onset and offset of each sentence constituent, pupil data were examined in two time windows. Figure 14 illustrates the time course of the averaged pupil dilation measurements with the average sentence constituent borders marked by vertical lines.

[6] Pupil size areas, averaged across trials and then across participants, showed that the monolingual group had larger pupil size areas (1397 in arbitrary units) than the bilingual group (959 in arbitrary units) and the Mann-Whitney-U-Test showed a significant difference ($W = 0$, $p < .001$).

Figure 14

Pupil dilation for SVO, OVS early and OVS late separated by group



Note. The dark gray area in the plot indicates time window 1 (1000-2000 ms) and the light gray indicates time window 2 (2000-3000 ms). The y-axis represents the averaged pupil dilation with respect to the baseline 0 in arbitrary units, and the x-axis illustrates 3500 ms time relative to sentence onset. Vertical lines indicate the means of the onset of the test sentence constituent (NP1 started with sentence onset by 0 ms, the onset of the verb had a mean of 750 ms and the onset of the NP2 a mean of 1250 ms). Error bars represent ±1 SE of between-subject variance.

It needs to be considered that ambient light conditions have an impact on pupil dilation (Ong et al., 2019): under bright light conditions, pupils constrict and under conditions of darkness, they dilate. In this study, monolingual children were tested in a laboratory room with constant artificial light and blinds down, whereas the bilingual children were tested in different kindergartens or at home in rooms with natural, potentially changing light. Hence, different lighting conditions during testing sessions may have led to variations in pupil dilation in the two groups. Consequently, group differences were not included in the analysis.

In this LMM, fixed effects were language group and condition; random effects were participants and items. Repeated contrast was used for condition: the contrast OVS early - SVO with positive estimates means that changes in pupil size were larger for the condition OVS early than for the condition SVO, and negative estimates means that changes in pupil size were larger for SVO than OVS early (SVO is subtracted from OVS early). In the contrast OVS late - OVS early, OVS early is subtracted from OVS late, i.e., positive estimates indicate larger pupil size changes for OVS late and negative estimates indicate larger pupil size changes for OVS early. In the following two LMMs, the data of the monolingual group were analyzed for time window 1 and time window 2 (see also Appendix C.18-C.19). Estimates, SE, df, *t*-values and *p*-values are reported in Table 14. Only significant results are discussed.

Table 14

Parameters for the LMMs for pupillometry with monolinguals in the intercept

Time window	Fixed effect	Estimate	SE	df	*t*-value	*p*-value
1	Intercept (grand mean)	98.097	8.805	61.037	11.141	<.001***
1	OVS early-SVO	-0.591	12.027	60.383	-0.049	=.961
1	OVS late-OVS early	10.983	13.888	60.375	0.791	=.432
2	Intercept (grand mean)	107.97	10.41	53.088	10.368	<.001***
2	OVS early-SVO	9.522	12.73	60.24	0.748	=.457
2	OVS late-OVS early	-0.457	14.70	60.23	-0.031	=.975

Within the monolingual group there was no significant difference in pupil dilation between conditions, neither in time window 1 nor in time window 2.

To check comparisons across conditions in the bilingual group, further analyses were run and reported, as displayed in Table 15 (see Appendix C.20-C.21 for model outputs). Simple effects and random effects were coded as for the previous model.

Table 15

Parameters for the LMMs for pupillometry in the bilingual group

Time window	Fixed effect	Estimate	SE	df	*t*-value	*p*-value
1	Intercept (grand mean)	66.47	14.51	55.97	4.582	<.001***
1	OVS early-SVO	-17.90	18.60	60.25	-0.963	=.340
1	OVS late-OVS early	44.79	21.48	60.25	2.085	=.041*
2	Intercept (grand mean)	75.552	17.271	55.098	4.374	<.001***
2	OVS early-SVO	4.755	21.847	60.248	0.218	=.828
2	OVS late-OVS early	24.184	25.228	60.251	0.959	=.342

* p <.05, *** p < .001

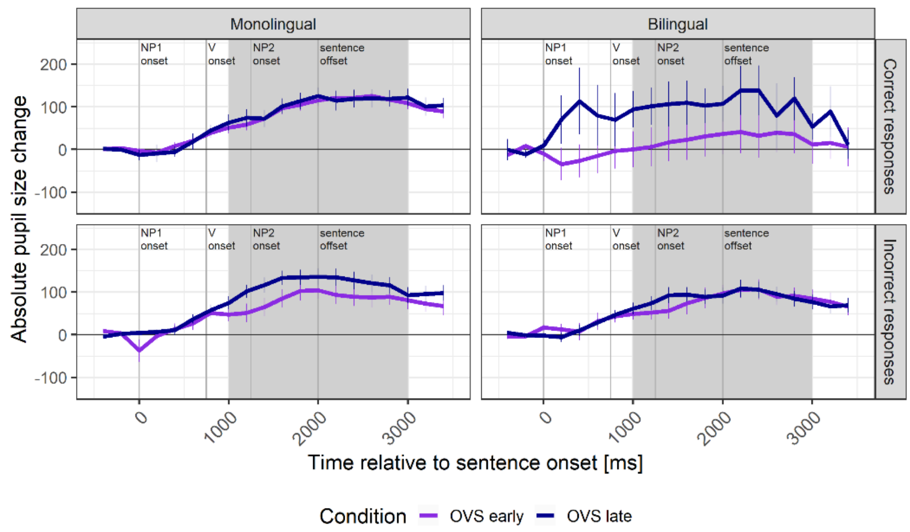
In time window 1, bilinguals' pupil dilation was larger for OVS late than for OVS early (p = .041). In time window 2, there was no difference in pupil size between conditions in the bilingual group.

In summary, a difference in pupil dilation between conditions was only observed in the bilingual group: during OVS late sentences, children had larger pupil dilation than during OVS early sentences.

However, this was only the case during time window 1 (note that the time window 1 was between 1000 and 2000 ms and that the onset of the NP2, namely the disambiguating cue for OVS late sentences, had a mean of onset of 1250 ms). Concerning participants' eye-gaze data, it was also investigated whether pupil dilation was related to the participants' responses. The aim of this analysis was to show whether cognitive effort represented by pupil dilation was reflected in incorrect or correct responses. Even with incorrect responses, participants' pupil dilation might signal that they have detected the disambiguating cue necessary to correctly assign thematic roles, but due to final inattention or working memory costs, they may give an incorrect answer on the sentence comprehension task. Thus, a difference between offline and online data may occur. In this case, the pupil data of incorrect responses may be similar to those of correct responses, or even higher, demonstrating a major processing effort during spoken language comprehension, independently of offline data. On the other hand, participants who gave incorrect responses may not have detected the disambiguating cue at all, and such pupil data may differ from those of correct responses. The pupil data included in the following analysis are those between 1000 ms and 3000 ms, meaning that time windows 1 and 2 from the previous analysis were combined. The SVO condition was removed from the models (for the same reasons discussed in Section 5.2.3.3). Figure 15 illustrates the time course of the averaged pupil dilation for correct and incorrect responses in both language groups.

Figure 15

Pupil dilation in OVS early and OVS late broken down by group and responses



Note. The dark gray area in the plot indicates the only time window between 1000 and 3000 ms. The y-axis represents the averaged pupil dilation with respect to the baseline 0 in arbitrary units and the x-axis illustrates 3500 ms time relative to sentence onset. Error bars represent ±1 SE of between-subject variance.

Two LMMs were run: the first included correct responses coded with treatment contrast in the intercept and the factor condition coded with repeated contrast. Monolinguals' data were analyzed.

Random effects were three random components: participants and items as random intercepts and the individual adjustment of condition per participant as random slope (see Table 16 and Appendix C.22 for complete model output).

Table 16

Parameters for the LMMs of monolinguals' pupil data with correct responses in the intercept

Fixed effect	**Estimate**	**SE**	**df**	***t*-value**	***p*-value**
Intercept (Corr. resp, grand mean)	105.7	12.13	42.30	8.721	<.001***
OVS late-OVS early (corr. resp)	-4.27	18.50	51.49	-0.231	=.818
Incorr. resp (corr. resp)	-1.23	0.568	425300	-2.162	=.031*
OVS late-OVS early x Incorr. resp	32.00	1.137	425000	28.128	<.001***

* p <.05, *** p < .001

The main effects for the monolingual group were that their pupil dilation did not show any differences between conditions in correct responses (p = .818) and also that their pupil dilation was smaller in incorrect responses than in correct responses across conditions (p = .031). Interaction effects showed that the pupil dilation difference between OVS late and OVS early in the monolingual group was larger in incorrect responses than in correct responses (p < .001).

The second model with bilinguals' data displays correct responses in the intercept coded with treatment contrast and the factor condition coded with repeated contrast (see Table 17 and Appendix C.23). Random effects were coded as in the previous model. Some effects of this models have not been discussed, since they reflect the results of the previous model.

Table 17

Parameters for the LMMs for pupil data with correct responses and bilinguals in the intercept

Fixed effect	**Estimate**	**SE**	**df**	***t*-value**	***p*-value**
Intercept (Corr. resp, grand mean)	66.014	22.93	49.981	2.880	=.006**
OVS late-OVS early (corr. resp)	79.031	38.73	56.675	2.041	=.046*
Incorr. resp (corr. resp)	18.697	1.398	419855	13.378	<.001***
OVS late-OVS early x incorr. resp	-68.381	2.795	419452	-24.465	<.001***

‡ p < .1, * p <.05, ** p <.01, *** p < .001

In the case of correct responses, bilinguals' pupil dilation was larger for OVS late than for OVS early (p = .046). Bilinguals' pupil dilation was larger in incorrect responses than in correct responses across conditions (p < .001).

In sum, the pupil dilation difference between OVS late and OVS early sentence is only present in the bilingual group and only when children gave correct responses. Monolinguals showed larger pupil dilation during the processing of both types of OVS sentences in which they gave correct responses than in incorrect responses. In the bilingual group, exactly the opposite pattern was found. The results of the bilingual group might indicate a detection of the disambiguating case-marking cues, also in the case of incorrect responses in the offline task.

5.2.4 Executive function tasks

Flanker task

In the Flanker task, the practice items as well as the test items with missing responses, response latencies longer than 4000 ms or shorter than 200 ms were excluded from the analysis. This affected 0.8% of the data. Monolingual children responded with 94.4% accuracy in the neutral condition, with 96.5% accuracy in the congruent condition and with 83.1% in the incongruent condition. The performance of bilingual children was similar. They responded correctly 96.4% of the time in the neutral condition, 95.6% of the time in the congruent and 83.7% of the time in the incongruent condition. In the analysis of RT, only correct responses were included so that 8.6% of the data were removed. Mean and median RTs for each condition, SD and range values for each group are visualized in ms in the following Table.

Table 18

Descriptive values for language groups in the Flanker tasks separated by conditions

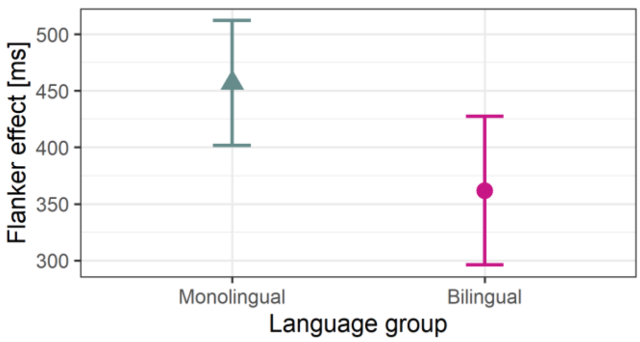
	Neutral	Congruent	Incongruent
Monolinguals			
Mean	1222	1229	1686
Median	1105	1163	1508
SD	292	275	479
Min	905	664	935
Max	2194	1793	2876
Bilinguals			
Mean	1349	1471	1833

	Neutral	Congruent	Incongruent
Median	1319	1392	1696
SD	300	361	450
Min	926	939	1197
Max	1970	2576	2802

Figure 16 shows the Flanker effect (RT of incongruent minus congruent trials) as a measure of inhibition in children. A better developed inhibition ability is reflected in the smaller Flanker effect.

Figure 16

Mean of the Flanker effect in both language groups



Note. Monolingual children are represented in gray and bilingual children in pink. Error bars represent ±1 SE of between-subject variance.

The Flanker effect in the monolingual group was 457 ms, while the bilingual it was 362 ms. The Shapiro-Wilk test revealed that the data were not normally distributed (W = .947, p = .01). Therefore, the non-parametric Mann-Whitney-U-Test was used to analyze whether there was a statistical difference between the Flanker effects in the two groups. The test did not reveal a significant difference between monolingual and bilingual children (W = 583, p = .149).[7]

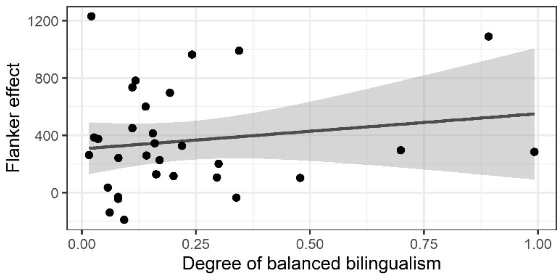
However, some studies (Bialystok & Majumder, 1998) have found cognitive advantages in more proficient, balanced bilinguals, since the competition between the two languages could be higher than in less balanced and less proficient bilinguals. Therefore, a subgroup of balanced and of unbalanced bilinguals was created. For each bilingual child, the degree of balanced bilingualism was calculated as the ratio of the German PPVT score to the Italian PPVT and the ratio of the German TROG score to the Italian TROG. A score of 1 indicated perfect balance. However, considering that it is almost never the

[7] Log-transformed data did not change this result.

case that bilinguals are perfectly balanced in different domains of a language, bilingual children with a ratio score of 1 ± 0.15 were considered as balanced (for a similar calculation see Bialystok and Barac, 2012). That is, scores above 1 indicate German dominance and below 1 Italian dominance (see Appendix C.24 for all individual data). Just one child out of 31 children obtained a score beyond the range of 1 ± 0.15 in the ratio of PPVT, whereas 15 children (out of 31 children) were within the range of 1 ± 0.15 in the ratio of TROG. The fact that only one child was in the balanced bilingual range, concerning the PPVT test, could be related to the different versions of the German and Italian PPVT (see Section 5.1.2.2 for test details). Furthermore, the TROG tested different domains of receptive grammar with increasing degrees of complexity, such as negation, prepositions, plural, comparatives, passives, relative clauses and OVS sentences. Therefore, it appears as a more reliable measure of balance than the PPVT in this study. To investigate the relationship between the Flanker effect and the degree of balanced bilingualism in children, children's ratio scores of TROG were converted[8] to a scale from 0 to 1, in which 0 represents the more balanced degree of bilingualism between German and Italian and 1 a language dominance in German or Italian. The following Figure represents the correlation graph between the Flanker effect and the degree of balanced bilingualism in children.

Figure 17

Correlation plot Flanker effect and degree of balanced bilingualism



Note. Each dot represents one child. The degree of balanced bilingualism is indicated on the x-axis: the more the values are closer to 0 the more a level of balance between the two languages. Unit measurements on the y-axis are in ms. The shaded area indicates ±1 SE of between-subject variance.

The Spearman non-parametric correlation test displayed no correlation between the Flanker effect and the degree of balanced bilingualism (*rho* = .090, *p* = .630)[9].

[8] The value of 1 was subtracted from all ratio scores and the results were converted to absolute values as their distance from 0.

[9] A group comparison between balanced and unbalanced bilinguals in the Flanker task was also calculated. The mean RT of the Flanker effect was 331 ms for balanced bilinguals and 391 ms for unbalanced bilinguals, the Mann-Whitney-U-Tests showed no difference between the two groups (*W* = 111, *p* = .722).

In conclusion, monolingual and bilingual children performed similarly during the Flanker task and a greater language balance had no effect on bilingual children's performance.

Task-switching paradigm

Practice items were also excluded from the analysis of the task-switching paradigm. As in the case of the Flanker task, test items with RT below 200 ms were excluded. This affected 0.04% of the data. Accuracy for the monolingual children in the non-switch blocks reached 92.6% (recall, only non-switch trials were presented in non-switch blocks), non-switch trials of the switch blocks reached 89.7% and switch trials of the switch blocks reached 87.2%. Bilinguals' data showed 93.1% of accuracy in the non-switch blocks, non-switch trials of the switch blocks reached 93% of accuracy and switch trials of the switch blocks reached 86.8%. For the RT analysis, only the correct responses were included (90% of the data). Mean and median RTs for each condition, SD and range values for each group are recorded in ms in the next Table.

Table 19

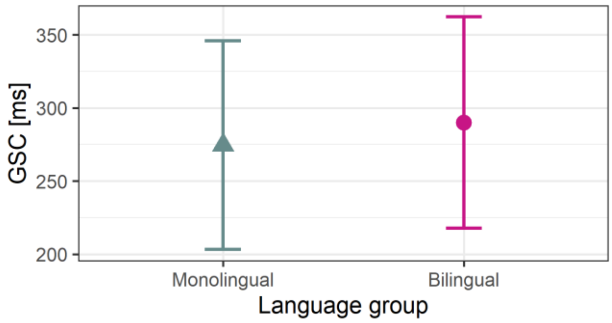
Descriptive values for language groups in the task-switching paradigm separated by conditions

	Non-switch blocks	Switch blocks non-switch trials	Switch blocks switch trials
Monolinguals			
Mean	1335	1610	1621
Median	1254	1441	1457
SD	536	577	622
Min	618	897	879
Max	3349	4594	5288
Bilinguals			
Mean	1333	1623	1634
Median	1257	1438	1467
SD	444	545	579
Min	670	941	954
Max	2764	3396	3798

Recall that global switch cost (GSC, RT of non-switch trials in switch blocks minus those of the non-switch trials in the non-switch blocks) was the variable used as a measurement for switching attention in children. Figure 18 reports the mean of GSC for monolingual and bilingual children.

Figure 18

Mean of the GSC in both language groups



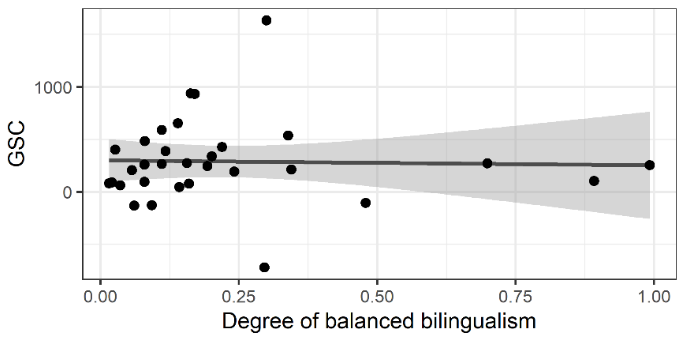
Note. Error bars represents ±1 SE of between-subject variance.

Monolinguals' GSC was 275 ms and bilinguals' GSC was 290 ms. Since data were not normally distributed (W = 896, p < .001), a Mann-Whitney-U-Tests was applied and no differences between the two groups were revealed (W = 491, p = .88).

As in the Flanker task, a correlation analysis between the GSC and the degree of balanced bilingualism of children was performed and is represented in the next Figure.

Figure 19

Correlation plot GSC and degree of balanced bilingualism



Note. Each dot represents one child. The shaded area indicates ±1 SE of between-subject variance.

Similarly as for the Flanker effect, the Spearman non-parametric correlation test showed no correlation between the GSC and the degree of balanced bilingualism (*rho* = .163, p = .381)[10].

[10] A group comparison between balanced and unbalanced bilinguals in the task-switching paradigm was also calculated. The mean RT of GSC was 225 ms for balanced and 351 ms for unbalanced bilinguals. However, no difference between the two groups was revealed by the Mann-Whitney-U-Tests (W = 96, p = .358).

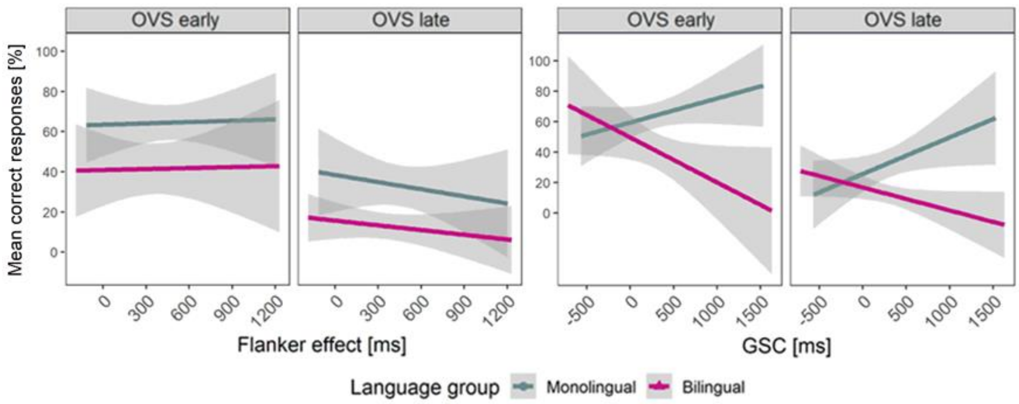
These results showed no differences between the performance of monolingual and bilingual children. Additionally, a higher language balance had no effect on bilingual children's performance, as was also the case in the Flanker task.

5.2.5 OVS sentence comprehension and executive functions

To investigate whether the two measurements for executive functioning (i.e., the Flanker effect and the GSC) were predictive of children's OVS sentence comprehension, LMMs were run. Regression lines are plotted in Figure 20 for monolinguals and bilinguals, showing how the mean proportion of correct responses differs as a result of varying EF skills.

Figure 20

Linear regressions reflecting the relation between mean percentage of correct responses and the Flanker effect and GSC separated by condition



Note. The x-axes report the time in ms for the Flanker effect and the GSC. Shaded areas indicate ±1 SE of between-subject variances.

For these analyses a repeated contrast was used for condition: OVS early was subtracted from OVS late, i.e., positive estimates indicated larger accuracy for OVS late, and negative estimates indicate larger accuracy for OVS early. Three random components were included in the models: two random intercepts for participants and items and a random slope, the effect of condition for each participant. The two variables: Flanker effect and GSC were centered to the groups average to make the intercept interpretable and added into the model as numeric predictors. Two LMMs were set up: in the first, monolingual children and the grand mean of correct responses in OVS sentences were in the intercept. The model's parameters are reported in Table 20 (see Appendix C.25 for complete model output).

Table 20

Results of the covariates analysis with the EF tasks with monolinguals in the intercept

Fixed effect	**Estimate**	**SE**	***z*-value**	***p*-value**
Intercept (Mono, grand mean)	-0.008	0.231	-0.034	=.973
Bili (grand mean)	-1.449	0.331	4.379	<.001***
OVS late-OVS early (Mono)	-1.635	0.334	-6.110	<.001***
Flanker effect (Mono)	-0.487	0.667	-0.282	=.778
GSC (Mono)	1.305	0.618	-2.427	=.015*
Bili x OVS late-OVS early	-0.403	0.408	-0.989	=.323
Bili x Flanker effect	-0.296	1.000	0.296	=.767
Bili x GSC	-2.811	0.852	-3.301	<.001***
OVS late-OVS early x Flanker effect	-0.625	0.873	-0.715	=.475
OVS late-OVS early x GSC	0.562	0.712	0.790	=.430
Bili x OVS late-OVS early x Flanker effect	-0.001	1.224	-0.001	=.999
Bili x OVS late-OVS early x GSC	-0.066	1.106	-0.059	=.953

* p <.05, *** p < .001

Fixed effects of conditions and groups are not discussed here. While the Flanker effect did not show a correlation with accuracy in sentence comprehension in monolinguals (p = .778), the GSC showed a positive correlation with accuracy. This means that an increase of GSC is related to an increase in accuracy (p = .015) across both OVS conditions in monolingual children. The interaction Bili x GSC shows that an increase of GSC is associated with a decrease of amount of correct responses, in comparison to the monolingual group (p < .001).

The second model included bilingual children and the grand mean of correct responses in OVS sentences in the intercept. The other parameters were set up as in the previous model. Table 21 reports these results (see Appendix C.26 for the model output).

Table 21

Results of the covariates analysis with the executive function tasks with bilinguals in the intercept

Fixed effect	**Estimate**	**SE**	***z*-value**	***p*-value**
Intercept (Bili, grand mean)	-1.456	0.249	-5.857	<.001***
Mono (grand mean)	1.449	0.331	4.379	<.001***
OVS late-OVS early (Bili)	-2.042	0.334	-6.110	<.001***
Flanker effect (Bili)	-0.188	0.667	-0.282	=.778
GSC (Bili)	-1.500	0.618	-2.427	=.015*
Mono x OVS late-OVS early	0.410	0.408	1.003	=.316
Mono x Flanker effect	-0.313	1.000	-0.313	=.754
Mono x GSC	2.808	0.851	3.299	<.001***
OVS late-OVS early x Flanker effect	-0.634	0.863	-0.735	=.463
OVS late-OVS early x GSC	0.506	0.844	0.600	=.548
Mono x OVS late-OVS early x Flanker effect	0.008	1.226	0.006	=.995
Mono x OVS late-OVS early x GSC	0.049	1.106	0.045	=.964

* p <.05, *** p < .001

In the bilingual group, a reduced GSC correlated to an increased number of correct responses (p = .015) across both OVS conditions. Just as in the monolingual group, the Flanker effect was not related to the number of correct responses in OVS sentence comprehension (p = .778).

To conclude, the lower the GSC in bilingual children, the more accurate they were in OVS sentence comprehension (negative regression line), while monolinguals demonstrated the opposite pattern: the lower the GSC, the less accurate they were in OVS sentence comprehension.

5.2.6 Linguistic, cognitive and social predictors

In this section, two different covariate analyses are reported. The first analysis investigated which predictors from the language background questionnaire were associated with the performance of bilingual (i.e., cumulative input and output, parents' education and the number of older school-aged siblings) and monolingual children (i.e., parents' education and the number of older school-aged siblings) in the sentence comprehension task. The second analysis concerned monolinguals' and bilinguals' data drawn from the standardized tests, in order to analyze to what extent language proficiency, memory capacity and reasoning ability affect the comprehension of German OVS sentences.

Before proceeding with the first covariate analysis for the bilingual group, correlations were calculated between the potential predictors derived from the language background questionnaire. The four variables were: children's German cumulative input and output, the parents' education and the number of children's older siblings. Spearman correlation coefficients are reported in the next Table.

Table 22

Correlation coefficients between variables from the language background questionnaire for bilinguals

	Input	**Output**	**Older siblings**	**Parents' edu**
Input	1			
Output	0.858***	1		
Older siblings	0.292	0.292	1	
Parents' edu	0.130	0.060	-0.047	1

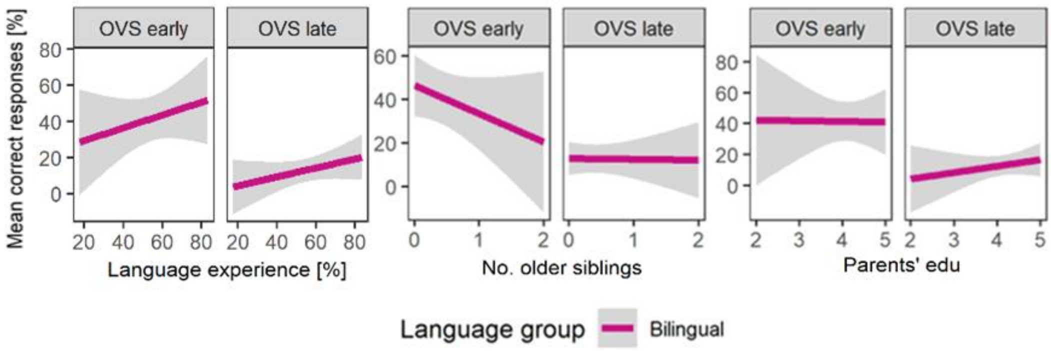
*** p < .001

Due to the high correlation between the predictor variables of language input and output, a new variable was created: language experience. This new variable is the mean of the variables cumulative input and output (see Section 5.2.1 for a reminder of how each variable was calculated).

Thus, Figure 21 linear regression lines were plotted for bilinguals, to reflect the predictors' language experience, number of older siblings and parents' education (x-axis) and the mean of correct responses (y-axis) separated by OVS early and OVS late sentences.

Figure 21

Linear regression reflecting the mean percentage of correct responses in the bilingual group for the effect of language experience, number of older siblings and parents' education



Note. The x-axis of language experience indicates a children's use of German (i.e., input and output) in percentages. The number of older siblings is indicated on the x-axis: 0 represents no older siblings, i.e., only children or children with only younger siblings. The x-axis of the parents' education shows a mean between the educational level of both parents on a scale from 1 to 5, in which 5 is the highest degree of education. Shaded areas indicate ±1 SE of between-subject variance.

LMM were run, in which a repeated contrast was used: OVS early was subtracted from OVS late, i.e., positive estimates indicated larger accuracy for OVS late and negative estimates indicate larger accuracy for OVS early. All continuous predictors were centered. Three random components were included in the models: two random intercepts for participants and items and a random slope, the effect of condition for each participant (see Appendix C.27). The three variables: language experience, number of older sibling (no. older siblings) and parents' education (parents' edu) were added into the model as numeric predictors. Table 23 reports these parameters.

Table 23

Covariates analysis from the language background questionnaire for bilinguals

Fixed effect	**Estimate**	**SE**	***z*-value**	***p*-value**
Intercept (grand mean)	-1.551	0.281	-5.512	<.001***
OVS late-OVS early	-2.123	0.374	-5.679	<.001***
Language experience	0.034	0.016	2.167	=.030*
No. older sibling	-0.390	0.405	-0.962	=.336
Parents' edu	0.197	0.435	0.452	=.651
OVS late-OVS early x lang. experience	0.016	0.020	0.789	=.430
OVS late-OVS early x no older siblings	0.928	0.497	1.867	=.062±

Fixed effect	Estimate	SE	*z*-value	*p*-value
OVS late-OVS early x parents' edu	0.937	0.590	1.588	=.112

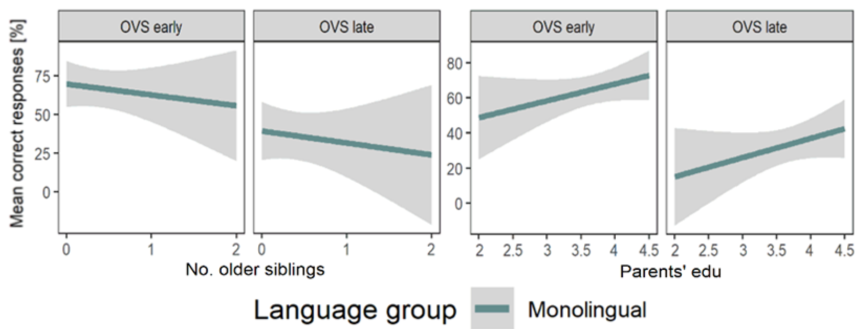
± p <. 01,* p <.05, *** p < .001

The language experience predictor was the only predictor that had a main effect on the comprehension of both OVS sentence conditions (p = .03) in bilinguals: children's language input (in the family and kindergarten including hours of current extracurricular activities, parents' reading aloud and the input of TV, music, and videogames) and language output was positively associated (positive estimate) with bilinguals' accuracy in their comprehension of OVS sentences. In other words, bilingual children who spent more time hearing and speaking German revealed a higher comprehension of OVS sentences.

Regarding the monolingual group, a similar analysis was performed to investigate whether there was an influence of older siblings or of the parents' education on their comprehension of OVS sentences (see Figure 22).

Figure 22

Linear regression lines reflecting the mean percentage of correct responses in the monolingual group for the effect of number of older siblings and parents' education



Note. The number of older siblings is indicated on the x-axis: 0 represents no older siblings, i.e., only children or children with only younger siblings. The x-axis of the parents' education shows a mean between the educational level of both parents on a scale from 1 to 5, in which 5 is the highest degree of education. Shaded areas indicate ±1 SE of between-subject variance.

The model specification was identical to the previous model, but without the language experience predictor (see Appendix C.28). Table 24 reports the parameters of this analysis.

Table 24

Covariates analysis from the language background questionnaire for monolinguals

Fixed effect	**Estimate**	**SE**	***z*-value**	***p*-value**
Intercept (grand mean)	0.101	0.318	0.319	=.750
OVS late-OVS early	-1.740	0.394	-4.421	<.001***
Parents' edu	0.896	0.505	1.774	=.076±
No older sibling	-0.498	0.524	-0.949	=.343
OVS late-OVS early x parents' edu	0.083	0.586	0.141	=.888
OVS late-OVS early x no older siblings	0.007	0.606	0.012	=.991

± p <.1, *** p < .001

Neither the parents' education nor the number of older siblings were predictors for the comprehension of OVS sentences in the monolingual group.

As mentioned at the beginning of this Section, the second covariate analysis was carried out to evaluate whether the performance of children in the comprehension of OVS sentences was modified by their linguistic and cognitive profiles. Spearman correlation coefficients were calculated for the raw scores of children in the standardized tests. These are reported in the following Table.

Table 25

Correlation coefficients between raw scores from the standardized tests

	German PPVT	**German TROG**	**Visual memory**	**Digit span**	**CPM**
German PPVT	1				
German TROG	0.635***	1			
Visual memory	0.500***	0.585***	1		
Digit span	0.306*	0.493***	0.542***	1	
CPM	0.187	0.293*	0.472***	0.351**	1

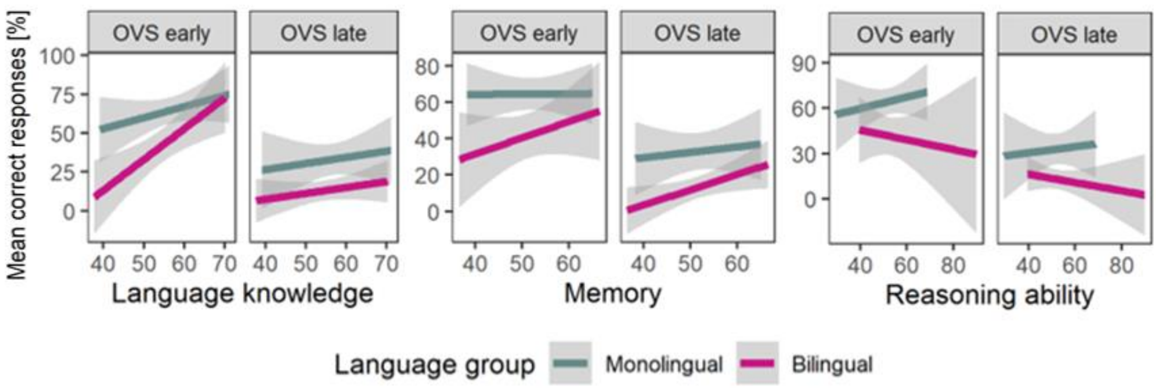
* p <.05, ** p <.01, *** p < .001

To reduce collinearity in the model, a composite variable for language knowledge (mean T-scores of the German PPVT and the German TROG) and for memory capacity (mean T-scores of the visual memory and digit span tests) were created. T-scores of the CPM were used as a reasoning ability measurement and included in the model as well.

Two levels of the factor condition (OVS early and OVS late) were compared by means of repeated contrast. The first analysis set the bilingual group in the intercept, and the second analysis the monolingual group. The three random components reflected the previous analysis. Figure 23 shows regression lines for bilinguals and monolinguals, representing the mean proportion of correct responses for the three potential predictors: language knowledge, memory and reasoning ability.

Figure 23

Linear regression lines reflecting the mean proportion of correct responses for the effect of language knowledge, memory and reasoning ability



Note. Gray lines represent monolinguals and pink lines bilinguals. Units of measurement on the x-axes are T-scores. Shaded areas indicate ±1 SE of between-subject variance.

Table 26 reports the parameters of the first model (see Appendix C.29 for complete model output).

Table 26

Covariates analysis with the standardized tests with bilinguals in the intercept

Fixed effect	**Estimate**	**SE**	***z*-value**	***p*-value**
Intercept (Bili, grand mean)	-1.437	0.255	-5.647	<.001***
Mono (grand mean)	1.371	0.331	4.136	<.001***
OVS late-OVS early (Bili)	-2.294	0.340	-6.747	<.001***
Lang knowledge (Bili)	0.053	0.029	1.815	=.070‡
Memory (Bili)	0.077	0.035	2.203	=.028*
Reas ability (Bili)	-0.054	0.028	-1.940	=.052‡
Mono x OVS late-OVS early	0.679	0.389	1.743	=.081‡

Mono x lang knowledge	-0.011	0.046	-0.235	=.814
Mono x memory	-0.097	0.050	-1.946	=.052‡
Mono x reas ability	0.076	0.039	1.957	=.050‡
OVS late-OVS early x lang knowledge (Bili)	-0.098	0.034	-2.860	=.004**
OVS late-OVS early x memory (Bili)	0.093	0.044	2.135	=.033*
OVS late-OVS early x reas ability (Bili)	-0.016	0.035	-0.455	=.649
Mono x OVS late-OVS early x lang knowledge	0.051	0.050	1.017	=.309
Mono x OVS late-OVS early x memory	-0.065	0.057	-1.135	=.256
Mono x OVS late-OVS early x reas ability	-0.010	0.045	-0.215	=.830

‡ *p* <.1, * *p* <.05, ** *p* <.01, *** *p* < .001

Fixed effects of conditions and groups are not discussed again, but reported for completeness. Across conditions, the predictor memory had a positive effect on the bilinguals' comprehension of both types of OVS sentences (*p* =.028). A significant interaction and negative estimate of language knowledge with condition indicated that the predictor was mainly present in OVS early (*p* =.004) and the positive estimate in the interaction OVS late-OVS early x memory indicated a stronger effect of memory in OVS late than OVS early sentences (*p* =.033). Table 27 records the parameters of the second model (see Appendix C.30 for complete output).

Table 27

Covariates analysis of the standardized tests with monolinguals in the intercept

Fixed effect	**Estimate**	**SE**	***z*-value**	***p*-value**
Intercept (Mono, grand mean)	-0.672	0.231	-0.291	=.770
Bili (grand mean)	-1.373	0.332	-4.136	<.001***
OVS late-OVS early (Mono)	-1.616	0.261	-6.187	<.001***
Lang knowledge (Mono)	0.042	0.035	1.189	=.235
Memory (Mono)	-0.019	0.036	-0.548	=.584
Reas ability (Mono)	0.022	0.027	0.821	=.412
Bili x OVS late-OVS early	-0.682	0.391	-1.745	=.081‡

Bili x lang knowledge	-0.011	0.046	0.233	=.816
Bili x memory	0.097	0.050	1.941	=.052‡
Bili x reas ability	-0.076	0.039	-1.957	=.050‡
OVS late-OVS early x lang knowledge (Mono)	-0.047	0.036	-1.301	=.193
OVS late-OVS early x memory (Mono)	0.029	0.037	0.775	=.438
OVS late-OVS early x reas ability (Mono)	-0.025	0.028	-0.892	=.373
Bili x OVS late-OVS early x lang knowledge	-0.051	0.050	-1.013	=.311
Bili x OVS late-OVS early x memory	0.065	0.057	1.136	=.256
Bili x OVS late-OVS early x reas ability	0.009	0.045	0.211	=.833

‡ p <.1, *** p < .001

Neither language knowledge nor memory and reasoning ability was associated to monolinguals' accuracy in their comprehension of OVS sentences (all p > .234), and no interactions were found (all p > .193).

To summarize, the exploratory analyses revealed that language experience and memory were a predictor for bilingual children, the higher the use of German (comprehension and production) and memory score, the better the comprehension of OVS sentences proved to be. None of the variables tested displayed any effects on the monolingual's comprehension of OVS sentences.

5.3 Discussion

The present study was primarily conducted to investigate the processing of German OVS sentences by German monolingual and German-Italian simultaneous bilingual children, and to explore whether this processing was influenced by EF abilities. Furthermore, language-external factors (such as the parents' education, the number of older siblings and language experience for bilingual children) and scores of receptive vocabulary and grammar, visual and short-term memory and reasoning ability were measured and correlated with children's accuracy in their comprehension of OVS sentences. Offline measures (i.e., accuracy of responses) as well as online measures (i.e., eye-gaze and pupillometry data), were collected and analyzed. OVS early sentences (i.e., disambiguating cue for accusative masculine nouns on the NP1) and OVS late sentences (i.e., disambiguating cue for nominative masculine nouns on the NP2) were created. Additionally, German weak and strong masculine nouns in the accusative case (i.e., with and without an added '-n' or '-en' as a suffix on the noun) were taken into account.

5.3.1 Sentence comprehension and processing

The RQ1 addressed whether OVS early sentences were easier to comprehend and process than OVS late sentences. The data revealed differences within and between the monolingual and bilingual group. The accuracy of monolingual children in their comprehension of OVS early sentences was above chance, while that of bilingual children was at chance level (i.e., 50% correct answers). It is suggested that German monolingual children seem to recognize the accusative case marking on the NP1, thereby overcoming the agent-first bias, whereas the chance level of bilingual children indicates that they find themselves in a linguistic phase behind that of monolingual children. Children's performance in OVS late sentences was below chance in both groups. The below chance level performance suggests that the agent role was assigned to the NP1. It seems to be that monolingual and bilingual children have difficulties in the integration of the disambiguating nominative cue on the NP2 in correctly assigning thematic roles, which supports the notion that children's difficulties in understanding sentences with late-arriving cues may be related to revision problems. Monolinguals outperformed bilinguals in their comprehension of both types of OVS sentences, and the performance by both language groups was higher for OVS early than for OVS late sentences.

The performance above chance by monolingual children during their comprehension of OVS early sentences is comparable with that of the 7-year-olds tested by Cristante (2016) and also that of the 7-year-olds tested by Dittmar et al. (2008). However, the German 6-year-olds tested by Biran and Ruigendijk (2015) and by Schipke et al. (2012) did not perform above chance. Concerning the experimental design and the sentence conditions (even though the sentences in Cristante (2016) were used in the past tense and not in the present), all these studies were performed in a similar way: children had to point to or name the correct picture, while two pictures were simultaneously presented with sentences that involved reversable theta roles. However, Biran & Ruigendijk (2015) also tested Wh-questions in the same experiment. This required children to switch between different types of non-canonical sentences, which may have made the task more difficult. Regarding the comprehension of OVS late sentences the performance of children in this study is comparable with children's performance reported by Cristante (2016) and Biran & Ruigendijk (2015), in which it was always below chance. It seems that the length of the ambiguous region affected children's comprehension: the longer the ambiguous region of the sentence was, the more difficulties children exhibited in reassigning a thematic role previously assigned. When the disambiguating cue appears on the NP1, sentence comprehension processing may be easier in comparison with sentences in which the disambiguating cue is on the NP2, maybe because there is more material intervening in between (see i.e., Bader et al., 2000; Choi & Trueswell, 2010; Ferreira & Henderson, 1998). The bilingual children tested in this study displayed the same level of accuracy in SVO sentences as the monolingual children did, suggesting that the monolingual advantage in sentence comprehension is not present for all linguistic phenomena. The Turkish-German eL2-learners tested by Cristante (2016) showed below chance level performance in both types of OVS sentences, while the simultaneous bilingual children in this study only performed below chance level in the OVS late sentence condition. It is thus suggested that children's understanding of OVS sentences is sensitive to length of exposure and AoO.

Eye-tracking data showed that German monolingual and German-Italian bilingual children started to look more to the target picture after the onset of the verb during the processing of SVO sentences than was the case in OVS early sentences. This means that the NP1 marked with the nominative masculine cue (*der*) leads children to look more at the correct picture than the accusative masculine cue (*den*) does. This difference in processing between SVO and OVS early sentences demonstrates that both monolingual and bilingual children were able to detect the disambiguating nominative case cue at the NP1. However, from sentence offset in OVS early sentences, monolingual children were above chance (meaning that their looking behavior was directed to the target picture for more than 50% of the time) suggesting that they waited for the disambiguating nominative cue to the NP2 in order to assign thematic roles. On the other hand, German-Italian bilingual children did not demonstrate a different looking behavior before and after the disambiguating cues in any of the time windows during the processing of OVS early sentences. Their looks to the target were at chance level during all time windows, suggesting a sensitivity to the difference between nominative and accusative case, even though they did not rely on the second disambiguating cue to the NP2 for a correct assignment of thematic roles. This is in line with the findings of Cristante (2016): 7-year-old eL2 German learners were not completely insensitive to the accusative case-marking cue on the NP1.

Monolingual and bilingual children showed more looks to the target during the processing of OVS late sentences from the NP2. The increase in looks to the target after having encountered the disambiguating nominative cue on the NP2 suggests that a revision process had started. It is argued that children detected this cue and noticed that its information was not in line with the previous thematic role assignment. Furthermore, in Sil3 after sentence offset (i.e., 2250 ms after sentence offset), monolingual children's target looking behavior was at chance level. This is a further indication that the disambiguating nominative cue at the NP2 had been detected. The only difference between monolingual and bilingual children was that bilinguals' looks to target remained below chance level in Sil3. This might indicate that monolingual children were more efficient in using the nominative case marker on NP2 than bilingual children. In Cristante (2016) neither L1 nor eL2 7-year-olds demonstrated any revision processes after the disambiguating cue at the NP2.

Pupillometry data revealed that the monolingual group did not exhibit any differences in their pupil size changes between the processing of SVO, OVS early and OVS late sentences, whereas bilingual children showed a larger pupil dilation for OVS late than they did for OVS early sentences during the first time window, namely between 1000 and 2000 ms. The reason why monolingual children did not show pupil size change during the processing of the different sentence conditions could be that they are at a different stage of case-marking acquisition and so do not require as much cognitive effort as bilingual children. On the other hand, the higher cognitive effort by the bilingual group in processing OVS late than OVS early sentences was detectable with a larger pupil size dilation in OVS late sentences.

The RQ2 asked whether OVS early sentences with weak masculine nouns were easier to comprehend than OVS early sentences with strong masculine nouns. Data have shown that neither monolingual nor bilingual children profited from the added suffix on accusative weak masculine nouns to assign thematic roles. It is suggested that the morpheme on weak nouns did not increase children’s performance in their comprehension of OVS sentences, presumably because the case-marking cue is already present in the determiner of the noun.

The RQ3 asked whether the comprehension and processing of OVS sentences was different between the monolingual and the bilingual groups. During the sentence comprehension task, bilingual children gave fewer correct answers than monolingual children in both types of OVS sentences. From sentence offset, bilingual children also displayed fewer looks to the target in comparison to monolingual children during both types of OVS sentences. These results indicate delay effects for comprehension and processing in the bilingual group, compared to the monolingual group. The fact that Italian does not assign case on the determiner but by means of a clitic might be one reason why bilingual and monolingual children display a different performance.

The RQ4 asked whether the performance of children was different during offline and online tasks. To answer this question, the factor accuracy (i.e., correct and incorrect responses) was added as variable into the LMMs of eye-gaze and pupillometry data. The eye-gaze data showed that both groups looked less to the target when giving incorrect responses than correct responses across both type of OVS sentences. This means that the eye-gaze data predicted children’s responses, and a different performance of children measured on accuracy and looking behavior was not observed. This is not in line with previous studies, that have shown how children’s performance in language comprehension tasks is higher in online than offline methods (Brandt-Kobele & Höhle, 2010; Höhle et al., 2016; Sekerina et al., 2004). However, pupillometry data showed a different pattern, though only in the bilingual group. Adding the factor accuracy into the model, bilinguals showed a larger pupil dilation in incorrect responses than in correct responses across conditions. This finding may suggest that children detected the disambiguating cues, but this was not sufficient to generate a correct interpretation of the sentence. Incorrect responses may be due to final inattention, offline task difficulties or working memory costs. Bahlmann et al. (2007) suggested that information during the processing of non-canonical sentences needs to be stored in the working memory. Testing a group of German adults in a fMRI study using the same two types of OVS sentences as in this study, Bahlmann et al. (2007) found more activation in the left inferior frontal cortex during the processing of OVS early sentences and an activation of the left supramarginal gyrus, a brain region involved in reanalysis aspects of sentence processing, during the processing of OVS late sentences. The activation of the left inferior frontal cortex was interpreted as a greater demand for working memory, caused by the complex syntactic structure, and the activation of the left supramarginal gyrus was associated with a reanalysis of the sentences initially interpreted using a subject-/agent-first strategy. Thus, it seems to be that working memory costs and reanalysis process are involved in processing OVS early and OVS late sentences respectively.

5.3.2 Sentence comprehension and executive functions

The RQ5 concerned the role that EF abilities may have on children's performance in sentence comprehension tasks. It was hypothesized that if bilingual children showed a cognitive advantage over their monolingual peers, then they might also perform better in the language comprehension tasks thanks to their better EFs.

To investigate this RQ, it was first checked how monolingual and bilingual children performed on the EF tasks. The Flanker effect (measure for inhibition ability) of the Flanker task and the GSC (global switch cost, measure for switching attention ability) of the task-switching paradigm were comparable in monolingual and bilingual children. Since some studies (e.g., Bialystok & Majumder, 1998; De Cat et al., 2018) have suggested that balanced bilingualism is more tightly connected to advanced EFs, I calculated a degree of balanced bilingualism in children by using their raw scores of the receptive grammar tests TROG (Bishop, 2003) in the German and Italian versions. It should be remembered that the first criterion used to account for balanced bilingualism was the selection of children from one parent-one language households. It was only in a second step that the degree of balanced bilingualism was calculated as the ratio of German TROG score to Italian TROG, a more finely grained measurement. However, the findings proved that the balanced bilingual children did not show any benefits when performing EF tasks over the unbalanced bilingual children. The way in which researchers have defined language balance differs a lot. Some studies determined language balance by basing it on children's mean length of utterance (see e.g., Cantone et al., 2006), others based it on parental language background questionnaires and the amount of children's input and output calculated from them (Torregrossa et al., 2021; Unsworth, 2015), while others again selected their samples according to the country where children had been born and to parental bilingual input (e.g., Argyri & Sorace, 2007), and others examined the language of instruction in schools (e.g., Bialystok & Majumder, 1998), whereas others define language balance by basing it on children's performance in both languages measured by standardized tests (e.g., Bialystok & Barac, 2012). This heterogeneity in measuring language balance makes the comparison between studies difficult and speaks in favor of the necessity to find parameters that systematically identify dominance in bilingualism. It cannot be excluded that the setting of other parameters to identify the degree of balanced bilingualism in this study might lead to differing results.

Thereafter, a correlation between the accuracy score of the sentence comprehension task and that of the EF tasks was calculated for all bilingual children. The Flanker effect did not influence children's OVS sentence comprehension, while the results for the task-switching paradigm showed varying correlations between the groups. The prediction was that higher accuracy scores should correlate with lower GSC in the task: the higher the accuracy, the lower the GSC should be. This correlation was only found in the bilingual group. The monolingual group showed the opposite pattern: higher GSC correlated with higher accuracy in the comprehension of OVS sentences. It might be that EF abilities tend to improve bilingual children's performance in their comprehension of OVS sentences since the linguistic structures necessary to resolve the task are not yet fully acquired and therefore difficult to master. The performance of monolingual children was above chance in the OVS early sentence condition and reached the chance level (as seen in the eye-tracking data) in the OVS late sentence condition. This might mean that they already have access to the linguistic competences necessary to reassign thematic roles and do not need the support of EFs.

It may be that the Flanker task is not the appropriate task to reveal a relationship between OVS sentence comprehension and the inhibition ability in 5;8-year-old children. With respect to the Flanker task, the task-switching paradigm is more difficult since it contains two tasks (i.e., task blocks in which only a single task appears and blocks where the tasks alternate) and a working memory component. Participants need to keep in mind the meaning of the task cue, in order to assign the stimuli to the target. Furthermore, the costs examined are associated with resolution of task-set interference from ambiguous stimuli and are believed to be involved in attention, planning and goal processing (e.g., Wiseheart et al., 2016). It might be that the EFs involved in the task-switching paradigm played a major role during the sentence comprehension task, in comparison to the EFs measured by means of the Flanker task.

5.3.3 Sentence comprehension and linguistic, cognitive and social predictors

The RQ6 asked whether language-external factors such as parents' education, the number of school-aged older siblings and, only for the bilingual group, the language experience in German played a role in children's OVS sentence comprehension. First, the parents' education and the number of older school-aged siblings were compared between the two groups. While the parents' education was higher in the bilingual than in the monolingual group, the number of older school-aged siblings was comparable in both groups. However, no variable affected the comprehension of OVS sentences, which holds for both groups. Even though previous studies (for the number of school-aged older siblings see Bridges & Hoff, 2014; for the parents' education see e.g., Calvo & Bialystok, 2014; V. C. M. Gathercole et al., 2016; Meir & Armon-lotem, 2017) found a benefit of these variables on children's language knowledge, this was not the case in this study. Nonetheless, the effect of parents' education and the number of school-aged older siblings was never previously investigated in association with the mastery of a specific linguistic phenomenon, such as the comprehension of German OVS sentences. Previous research reported the existence of benefits of these variables on more general language knowledge, such as vocabulary and grammar measured by standardized tests. In the bilingual group, the language experience in German was a relevant predictor for OVS sentence comprehension, i.e., the more the children used German, the higher the accuracy in their comprehension of OVS sentences was. This was expected and in line with previous studies (see e.g., Unsworth, 2015).

The RQ7 aims at investigating whether children's comprehension of OVS sentences was affected by general linguistic and cognitive measurements. Both groups were first compared in the performed standardized tests: receptive vocabulary, receptive grammar, visual memory, short-term memory, and reasoning ability. Their performance was comparable in all tests. Composite variables were then calculated on the basis of correlation effects and test similarities: language knowledge, memory and reasoning ability. In the monolingual group, none of these factors influenced children's comprehension of OVS sentences, whereas in the bilingual group the analyses showed that memory had a positive effect on both OVS sentences (even stronger for OVS late sentences). This finding is in line with previous results which have suggested that information needs to be stored in working memory during the comprehension of complex sentences (see e.g., Bahlmann et al., 2007; S. E. Gathercole et al., 2004).

5.3.4 Summary

Through this study it has been shown that, at the age of 5;8, monolingual German children still have some problems in using case-marking to assign thematic roles. Compared to monolingual children, simultaneous German-Italian bilingual children exhibited more difficulties during the comprehension and processing of OVS sentences. For both groups the length of the ambiguous region played a role: when the disambiguating cue appeared late in the sentence, children's performance was lower than when the disambiguating cue appeared early in the sentence. This means that, the integration of the disambiguating cue into the processing is more difficult when the cue occurs late in the sentence. The delay effects shown by the bilingual group in comparison to the monolingual group in their comprehension and processing of OVS sentences might indicate cross-linguistic influences. This may mainly be due to the fact that the German and Italian languages have different case-marking systems and therefore require diverging cues to assign thematic roles. It was also investigated whether EF abilities, such as inhibition and switching attention ability, played a role in children's comprehension of OVS sentences. These findings showed that the inhibition ability did not play a role in children's comprehension of OVS sentences, whilst the switching attention ability had a different influence in the two groups: bilingual children with lower GSC showed a better comprehension of OVS sentences, while monolingual children with lower GSC showed a lower comprehension of OVS sentence comprehension. The bilingual children's processing of German OVS sentences was affected by the amount of German use with and by the children and by their memory capacity. It is suggested that cognitive factors lead to a better linguistic performance if the linguistic structure is not fully mastered.

6 Study 2: German passive sentence processing by L1 and 2L1 4-year-olds

Study 2 examines whether German monolingual and German-Italian simultaneous bilingual children are capable of using morphosyntactic cues needed to assign thematic roles during the comprehension and processing of eventive long passive structures with animate NPs. Children aged between 4 and 5 were tested by means of an offline task, namely a sentence-picture verification task, during an online eye-tracking experiment. Eye-gaze data and pupil data were collected and analyzed in order to investigate how children process the morphosyntactic cues of the German passive construction (i.e., auxiliary and the *by*-phrase) in real-time. German, long passive sentences with the auxiliary *werden* were used in the past tense form to avoid the temporary ambiguous reading given by the present tense form (also used to construct future tense sentences) even though this form *wurde* also overlaps with its use as a copula verb, e.g., *Die Frau wurde schwanger* 'The woman became pregnant' (see e.g., Abbot-Smith & Behrens, 2006; Knoeferle et al., 2005).

Two types of passive sentences were considered: passive sentences with the subject in the prefield before the finite verb (PAS PF, henceforth) and passive sentences with the subject in the middlefield after the finite verb (PAS MF, henceforth). PAS PF sentences had two constituents in the middlefield: a temporal adverb as first constituent and the *by*-phrase as second constituent (6.1).

(6.1) *Das Nilpferd wurde gestern von dem Panther geschoben*
The hippopotamus was yesterday by the panther pushed
'Yesterday the hippopotamus was pushed by the panther'.

PAS MF sentences had a temporal adverb in the prefield position, the subject as first constituent of the middlefield and the *by*-phrase as second constituent (6.2).

(6.2) *Gestern wurde das Nilpferd von dem Panther geschoben*
Yesterday was the hippopotamus by the panther pushed
'Yesterday the hippopotamus was pushed by the panther'.

As mentioned in Section 2.3, the Italian parallel structure for (6.1) – considering the S(finite)V alignment – is possible (see 6.3) whereas the Italian parallel construction for (6.2) is ungrammatical (see 6.4) since the subject must remain in a preverbal position to maintain the SVO order required in Italian when an adjunct is fronted in broad focus contexts (see 6.5).

(6.3) *L'ippopotamo venne spinto ieri dalla pantera*

The hippopotamus was pushed yesterday by the panther
'Yesterday the hippopotamus was pushed by the panther'.

(6.4) **Ieri venne spinto l'ippopotamo dalla pantera*
Yesterday was pushed the hippopotamus by the panther
'Yesterday the hippopotamus was pushed by the panther'.

(6.5) *Ieri l'ippopotamo venne spinto dalla pantera*
'Yesterday the hippopotamus was pushed by the panther'.

In passive sentences, the canonical mapping of subject/agent and object/patient is not maintained: the subject is the patient, and the referent in the optional PP has the role of agent. Passives may be difficult to parse due to this mismatch of syntactic functions and thematic roles. Thus, children need to override the agent-first bias to correctly interpret the NP1 as the patient of the sentence.

In German, the use of the subject in the pre- or middlefield leads to different hypotheses. Following the proposal of Huang et al. (2013), children should perform better when the passive voice cue appears before the NP1, preventing the assignment of the agent role to the NP1. This means that the children in this study should perform better in PAS MF than in PAS PF sentences. However, considering that German declarative sentences start with a subject more often than with an adverb (Bohnacker, 2007; Bohnacker & Rosén, 2008; Engel, 1974), PAS PF sentences may be easier to comprehend. Hence, the first RQ is:

RQ1: Do monolingual and bilingual children show a different performance in PAS PF sentences to PAS MF sentences?

If the adverb in the prefield prevents children from interpreting the NP1 as agent of the sentence, higher accuracy scores, more looks to the target picture and smaller pupil dilation are expected during the processing of PAS MF in contrast to PAS PF sentences. On the other hand, if the higher frequency of German declarative sentences with the sentence initial subject plays a role in children's performance higher accuracy scores, proportion of looks to the target and smaller pupil dilations are expected during the processing of PAS PF sentences.

With respect to German-Italian bilingual acquisition, it should be considered that the syntactic structure of PAS PF sentences overlaps in the two languages with regard to the S(finite)V alignment, whereas the structure of PAS MF sentences does not. Thus, the next RQ is formulated as follows:

RQ2: Does the bilingual group reveal better comprehension and faster processing of PAS PF sentences in comparison to the monolingual group?

According to N. Müller and Hulk (2000) the structural overlapping at the surface level of two languages is a criterion in which cross-linguistic influence can occur, and indeed there is an overlap in German and Italian, considering the S(finite)V alignment in PAS PF sentences. Thus, it is hypothesized that, during the comprehension and processing of PAS PF sentences (and not PAS MF sentences), bilingual children show similar accuracy scores and processing costs, or even better comprehension and faster processing, compared to monolingual children.

In order to investigate whether different methods display differences in children's performance on language comprehension, the following RQ was formulated:

RQ3: Is there a difference in children's performance between online and offline tasks?

As previous studies have demonstrated (e.g., Höhle et al., 2016; Sekerina et al., 2004), children's performance in online methods may be a better method of giving evidence of their linguistic competence than offline methods.

Additionally, as in Study 1, the role that EFs may have on children's sentence comprehension was considered. Monolinguals and bilinguals were tested by means of two EF tasks which assessed their inhibition and switching ability, thus the formulation of the next RQ:

RQ4: Are children's EFs related to their comprehension of German passive sentences?

It is hypothesized that higher accuracy scores in passive comprehension should correlate with children's performance in EF tasks. Better inhibition and switching attention abilities might help children in the reassignment of thematic role during their comprehension of passive sentences if the linguistic structure is difficult to master.

Additionally, if the bilingual group shows a cognitive advantage over their monolingual counterparts, it may also be the case that the performance of bilinguals in the sentence comprehension task is even higher than that of monolinguals: bilinguals may be facilitated by stronger EF abilities and show higher accuracy scores in the sentence comprehension task compared to monolinguals.

However, the linguistic performance of bilingual children may also be influenced by other factors. Their language experience, parents' education and the number of school-aged older siblings were taken into consideration as potential predictors of language comprehension in the bilingual group. The variables related to parents' education and number of school-aged older children were considered for monolingual children, too. Hence, the next RQ is:

RQ5: Do language-external factors predict children's comprehension of German passive sentences?

It is predicted that bilingual children's higher accuracy scores in passive comprehension should correlate with higher language experience in German. Higher parents' education and a greater number of school-aged older siblings are hypothesized to be predictors for the performance of both monolingual and bilingual group on passive comprehension.

Finally, children's receptive vocabulary and grammar, verbal and visual short-term memory and reasoning ability were assessed. It is possible that children's performance in passive sentences is influenced by these predictors. Therefore, the final RQ is formulated as follows:

RQ6: Is the comprehension of German passive sentences in L1 and 2L1 children influenced by their language knowledge, memory and reasoning skills?

A higher accuracy score in passive comprehension is supposed to correlate with higher accuracy in these three variables.

This chapter first describes the participants and the materials used to conduct the experiments, and then analyses the collected data. Finally, the discussion section answers the RQs.

6.1 Method

6.1.1 Participants

34 German monolingual children and 38 German-Italian bilingual children participated in this study. Selection criteria for bilingual children were AoO from birth in German and Italian and families using the one parent-one language approach. Two children with only Italian input at home had a minimum of *z*-scores of -0.4 (i.e., 34.5th percentile) in the German vocabulary and grammar tests, so consequently they were also included in the analysis. Data from 12 bilingual children were excluded for the following reasons: growth disorders (n=2), no understanding of Italian (n=2), missing the second test session (n=2) and dropouts due to boredom and ending test participation (n=6). Eight monolingual children were excluded after missing the second test session. In order to achieve the same sample size in both groups, children were matched by age. Consequently, four other monolingual children were excluded from the analysis of this study.

The final sample included 26 monolinguals (mean age: 4;6, range: 4;1-5;0, SD: 0;3; 10 girls) and 26 bilinguals (mean age: 4;5, range: 4;0-4;10, SD: 0;3; 18 girls). In three families, mothers had been raised as bilinguals (AoO from birth in German and Italian): one mother spoke both languages to the child, while the other two mothers almost always spoke Italian. In 11 families, the father was a native speaker of German and the mother a native speaker of Italian. In 10 other families, the father was a native speaker of Italian and the mother of German. The two remaining families are the two families named above, in which there were only Italian native speakers at home, but the two children went to a German monolingual kindergarten, just like seven other children. The rest of them, 17 children, went to different German-Italian bilingual kindergartens in Berlin.

According to the language background questionnaire completed by parents, none of the children had developmental disorders. Monolingual children were tested at the BabyLAB of the University of Potsdam, and bilinguals were either tested at home or at kindergarten, except for two bilingual children who also went to the BabyLAB. Children received a small book as a gift after the testing sessions, and parents visiting the lab additionally obtained a reimbursement for their travel costs. The Ethics Committees of the University of Potsdam approved this study, and parents signed a written consent form.

6.1.2 Materials and Procedure

6.1.2.1 Language exposure and socio-economic profile

The language background questionnaires used for bilinguals (see Appendix A.1) and for monolinguals (see Appendix A.2) were identical to those of Study 1. The used variables are reported in Section 6.2.1.

6.1.2.2 Standard assessments

Linguistic profile

As in Study 1, the German and Italian versions of the PPVT (Dunn & Dunn, 2007) were used to assess receptive vocabulary, and the German and Italian versions of the TROG (TROG; Bishop, 2003) to assess receptive grammar (see Section 5.1.2.2 for more details).

Cognitive profile

The CPM (Raven, 2003), word order (hereafter visual memory) and number recall (hereafter digit span) subtests of the K-ABC II (Kaufman et al., 2003) were administered to assess reasoning ability, visual working memory and memory span in monolingual and bilingual children (see Section 5.1.2.2 for more details).

6.1.2.3 Sentence comprehension and processing

For this study, 32 sentences were created. Each sentence described a scene between two animal characters: the agent and the patient. Eight animal labels were neuter nouns, and the other eight were masculine nouns (see Table 28).

Table 28

Animal labels used for Study 2

Neuter nouns	**Masculine nouns**
das Nilpferd ('hippopotamus')	*der Panther* ('panther')
das Schaf ('sheep')	*der Dachs* ('badger')
das Eichhörnchen ('squirrel')	*der Frosch* ('frog')
das Zebra ('zebra')	*der Hai* ('shark')
das Schwein ('pig')	*der Pfau* ('peacock')
das Känguru ('kangaroo')	*der Esel* ('donkey')
das Reh ('deer')	*der Flamingo* ('flamingo')
das Pferd ('horse')	*der Tiger ('tiger')*

The same two animal characters were always mentioned together in a sentence, creating a pair (each row in Table 28 represents a pair). Masculine and neuter nouns were chosen because they have the same article form in the dative case, that is used in the passive *by*-phrase (*von*-phrase), namely *von dem* ('by the'). Eight of the 32 sentences were used for the active condition with the subject NP in the prefield position (henceforth ACT PF, 6.1), 8 sentences were used for the active condition with the subject NP in the middlefield position (henceforth ACT MF, 6.2), 8 sentences were used for the PAS PF condition (6.3) and 8 sentences were used for the PAS MF condition (6.4).

(6.1) ACT PF
Der *Panther hat kürzlich das Nilpferd getreten*
[The panther]NOM has recently [the hippopotamus]ACC kicked
'The panther kicked the hippopotamus recently'

(6.2) ACT MF
Gestern hat der Panther das Nilpferd gefangen
Yesterday has [the panther]NOM [the hippopotamus]ACC caught

'The panther caught the hippopotamus yesterday'

(6.3) PAS PF

Der Panther wurde täglich von dem Nilpferd gekitzelt

[The panther]$_{\text{NOM}}$ was daily by [the hippopotamus]$_{\text{DAT}}$ tickled

'The panther was tickled by the hippopotamus daily'

(6.4) PAS MF

Einmal wurde der Panther von dem Nilpferd gewaschen

Once was [the panther]$_{\text{NOM}}$ by [the hippopotamus]$_{\text{DAT}}$ washed

'The panther was washed by the hippopotamus once'

Altogether, eight transitive verbs were used, and they always appeared in combination with the same adverbs. Table 29 reports them.

Table 29

Verbs and adverbs used for Study 2

Transitive verbs	**Adverbs**
fangen ('catch')	*gestern* ('yesterday')
treten ('kick')	*kürzlich* ('recently')
waschen ('wash')	*einmal* ('once')
kitzeln ('tickle')	*täglich* ('daily')
schlagen ('hit')	*damals* ('then')
messen ('measure')	*neulich* ('the other day')
schieben ('push')	*vorhin* ('just now')
rufen ('call')	*letztens* ('lately')

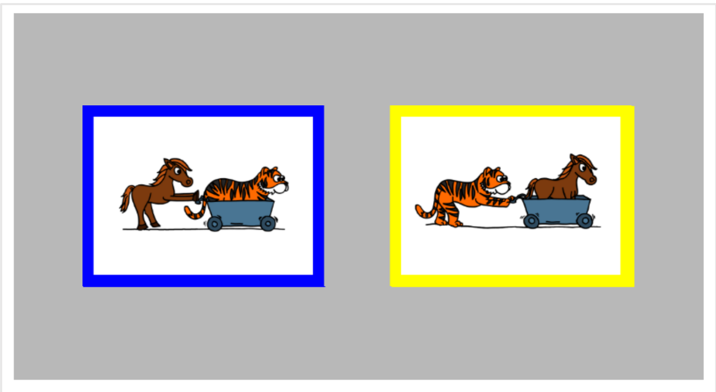
Six further sentences were created for the practice phase: two items contained a single word and four items, forming intransitive sentences. The experiment was divided into five blocks with short pauses in between and started with the practice phase. Eight items were included in each of the other four blocks: two items were presented in the ACT PF condition, two in the ACT MF condition, two in the PAS PF and two in the PAS MF condition. Each verb/adverb pair was used once per block. The character used as the subject in one condition also remained the subject for the other three conditions, but carried a different thematic role (e.g., in the examples (6.1)-(6.4) *the panther* is the subject with the role of agent in the active sentences and with the role of patient in the passive sentences). Each animal pair was used four times, once per sentence condition. Sentences were on average 2750 ms long: the mean length of the prefield (PF; NP1 or adverb) was 550 ms, of AUX (voice cue *hat* 'has' or *wurde* 'was') 250 ms, of the first constituent in the middlefield (MF1; NP1 or adverb) 550 ms, of the second constituent in the middlefield (MF2; NP2 for active sentences and PP for passive sentences) 800 ms and of the lexical verb (V) 600 ms. Four different versions of the experiment were created, in which each sentence occurred in a different condition in each version (see Appendix D for all versions). Each version was also presented in a forward and a backward order to ensure a varying item order.

As in Study 1, a female native speaker of German recorded the attention getter *Schau mal* ('Look at that') and the introductory question *Welche Farbe hat der Rahmen* ('What color is the frame'). The recordings were played during the first four practice items, in order to familiarize the children with the task. Thereafter, the introductory question was not asked again (except to elicit a participant answer after a few seconds' pause).

In total, 64 different pictures were created. For each sentence, two pictures, each showing two animals involved in an action, were presented on the eye-tracking monitor screen (1920x1080 pixels in size). The two pictures of each pair only differed with respect to the thematic roles of the two animal characters. Figure 24 represents the visual material for the item: *Das Pferd wurde gestern von dem Tiger geschoben*, 'The horse was pushed by the tiger yesterday'.

Figure 24

Example of visual stimuli array used in the passive experiment



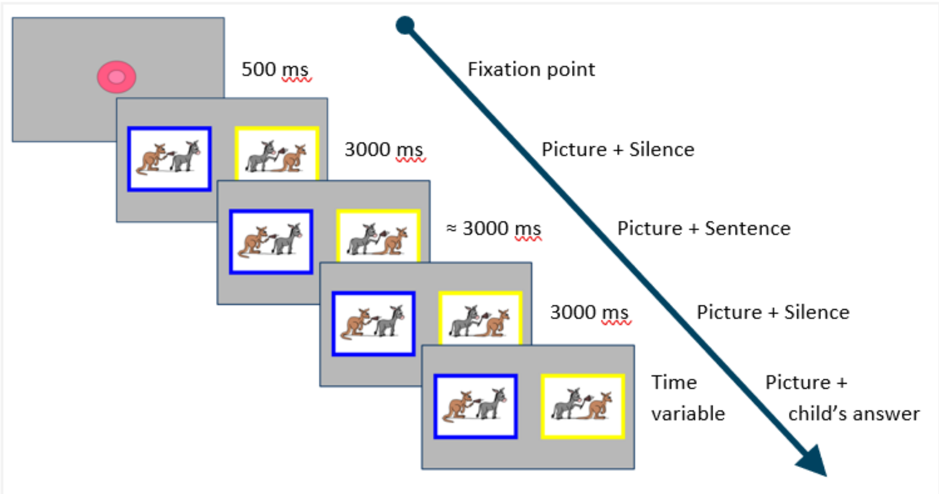
As in Study 1, the picture presented on the left of the screen always had a blue frame, while the picture on the right was framed in yellow. The color of each frame was 25x25 pixels per size and the space between the two pictures was 150 pixels in size. All the animal characters depicted were comparable in size and colored with a free graphic toll. The background color within the pictures was white and that of the rest the screen was gray (rgb color values: 184, 184, 184). In half of the displays, the animal characters were pointing to the right and in the other half to the left of the monitor. The two pictures (each 500x354 pixels in size) were presented simultaneously in the center of the eye-tracking monitor.

Procedure

The procedure was the same as used for Study 1 (see Section 5.1.2.3). The cover story of the eye-tracking experiment was introduced by a girl, Christina, who wanted to play a game with the child. Children were told that Christina showed pictures and said something about them; the children should decide which picture Christina was talking about, by naming the corresponding frame color, namely blue or yellow. Figure 25 illustrates the entire trial structure of the experiment in Study 2.

Figure 25

Experimental trial of the passive experiment



The trial started with a fixation point of (500 ms). The two pictures were presented for around 9000 ms: a static presentation of the pictures in silence preceded (3000 ms) and followed (3000 ms) the presentation of the pictures together with the test sentence (mean of duration of test sentences: 2711 ms, range: 2211 ms – 3272 ms). Children had optional time to label the color of the frame, after which the experimenter started the next trial manually. The 38 trials (32 test trials and 6 practice trials) were divided into five different blocks with short pauses between each one, in which the girl, Christina, appeared again on the screen, closer each time to the checkered flag, so that the children were encouraged to do well in finishing the game. The experiment was monitored for expected answers (half blue and half yellow), including no more than three consecutive same answers. The eye-tracking experiment lasted approximately 25 minutes.

6.1.2.4 Executive function tasks

The Flanker task (Eriksen & Eriksen, 1974; Rueda et al., 2004) and the task-switching paradigm (Jersild, 1927; Wiseheart et al., 2016) were used as EF tasks. See Section 5.1.2.4 for a detailed description of these tasks.

6.1.2.5 Study sessions

The course of the sessions was the same as in Study 1. Children were tested in two sessions around two weeks apart. Each testing session lasted 60 minutes for bilingual children and for monolingual children 45 minutes. As in Study 1, bilingual children were additionally tested by means of the Italian PPVT and the Italian TROG. Besides these two, all further tests were conducted in German.

6.1.3 Data analysis

The software used and the models run were the same as in Study 1 (see Section 5.1.3 for more information). As in Study 1, eye-tracking data were filtered by the data analysis software SR Research Data Viewer 4.1.1 (2019) and uploaded on to Rstudio. Two spatial AoI were defined. Each AoI was 500 x 354 pixels, corresponding to the size of the pictures without the frame. Proportions of looks were collapsed within participants and items for each time window. A time window was defined according to the mean length of the PF (550 ms), AUX (250 ms), MF1 (550 ms), MF2 (800 ms) and V (600 ms). Analogously to Study 1, three different time windows (Sil1, Sil2, Sil3) were considered after sentence offset, which all included 750 ms of silence (see also Section 6.1.2.3).

Correct responses in the sentence comprehension task were coded as 1, and incorrect responses as 0. The independent variables for the offline and online methods were sentence condition (ACT PF/ ACT MF/PAS PF/PAS MF) and language group (monolinguals/bilinguals). The dependent variables were the accuracy of correct responses and the proportion of looks to the target picture across the different time windows.

6.2 Results

6.2.1 Language exposure and socio-economic profile

The average proportion of cumulative input and output and the parents' education are plotted with mean, SD, median, min and max (i.e., range) values in Table 30.

Table 30

Covariates from the language background questionnaire for the 4-year-old bilingual group

	Mean	**SD**	**Median**	**Min**	**Max**
Proportion cumulative input in German (%)	58.4	10.9	57.3	40	85.4
Proportion cumulative output in German (%)	49.8	17.1	47.2	18.7	79.5
Parents' education	4.1	0.6	4	2	5

At the group level, children displayed a balanced exposure to German and Italian (mean: 58.4% for German and 41.6% for Italian) and a balanced language use (mean: 49.8% for German and 50.2% for Italian). No children received less than 40% input in German and no children spoke German for less than 18.7% of the time. The level of education of the bilinguals' parents was relatively high, at a mean value of 4.1, on a scale from 1 to 5. The parents' education was also calculated for the monolingual group (mean: 3.2; SD: 0.8; median: 3; range: 1.5-5). A correlation between mothers' and fathers' education in the monolingual group was found (*rho* = .505, *p* = .010), but this correlation was absent in the bilingual group (*rho* = .160, *p* = .434). However, the variable parents' education (i.e., the mean between mothers' and fathers' education) was used as predictor for both groups in further data analyses.

The number of older siblings was calculated for the bilingual group (mean: 0.8; SD: 0.5; median: 1; range: 0-2) and monolingual group (mean: 0.4; SD: 0.8; median: 0; range: 0-3). The sample of 26 bilingual children included five missing data, six only children, three children with younger siblings (two of them with one younger sibling and the other with two younger siblings) and 12 children with one older sibling. The sample of 26 monolingual children included missing data for seven children, 10 only children, five children with one younger sibling, and four children with older siblings (two children with one older sibling, one child with two older siblings, and one child with three older siblings). All older siblings were school-aged.

Additionally, the parents' education and the number of older siblings were compared between the monolingual and the bilingual groups. The Mann-Whitney-U-test showed a higher level of parents' education in the bilingual families than in the monolingual families (*W* = 524, *p* = <.001), and the bilingual children also had a higher average of older siblings than the monolingual children (*W* = 286.5, *p* = .008). These variables are part of the covariates analysis in Section 6.2.6.

6.2.2 Standard assessments

To test whether data were normally distributed the Shapiro-Wilk test was adopted. The results were that German PPVT (W = .986, p = .781), German TROG (W = .977, p = .407) and digit span (W = .956, p = .075) followed a normal distribution, while the data of visual memory (W = .905, p < .001) and CPM (W = .953, p = .045) were not normally distributed. As in Study 1, t-tests were used for normally distributed data and the Mann-Whitney-U-Tests for not normally distributed data. Table 31 shows descriptive data of the raw scores for all the standardized tests.

Table 31

Performance of 4-year-olds in the standardized tests and group comparisons

		Mean	SD	Median	Min	Max
PPVT (max. score 228)						
	Monolingual children	111	18.6	113	65	140
	Bilingual children	90	20.5	87	59	158
TROG (max. score 84)						
	Monolingual children	56.2	12.8	58	29	77
	Bilingual children	55.0	10.2	57	41	75
Digit span (K-ABC II) (max. score 21)						
	Monolingual children	6.8	1.8	7	3	11
	Bilingual children	7.6	2.3	7	3	12
Visual memory (K-ABC II) (max. score 31)						
	Monolingual children	10.5	2.8	11	5	16
	Bilingual children	10.6	4.2	10	5	25
CPM (max. score 36)						
	Monolingual children	13.7	2.9	14	10	19
	Bilingual children	15.6	3.1	16	10	24

A statistical difference between the two groups is present in the German PPVT ($t(50) = 3.950$, $p < .001$), in which the monolingual group outperformed bilingual children, and in the CPM, in which the bilingual group outperformed the monolingual group ($W = 435.5$, $p = .036$). The performance of monolingual and bilingual children proved to be comparable in the German TROG ($t(48) = 0.395$, $p = .695$), the digit span ($t(46) = -1.400$, $p = .168$) and the visual memory test ($W = 318.5$, $p = .720$).

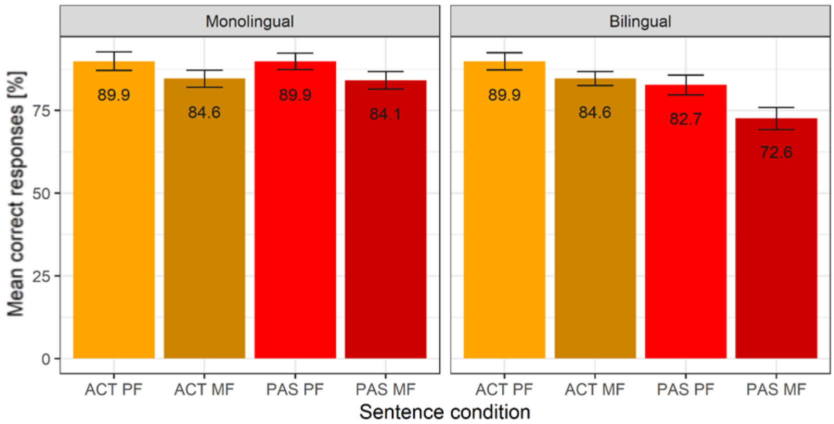
6.2.3 Sentence comprehension and processing

6.2.3.1 Accuracy

Mean percentage of correct responses of children in the sentence comprehension task, separated by condition and group, is shown in Figure 26.

Figure 26

Mean percentage of correct responses for 4-year-olds



Note. Monolingual children are represented in the left panel and bilingual children on the right. The four passive sentence conditions are reported on the x-axis and children's mean percentage of correct responses on the y-axis. Error bars represent ±1 SE of between-subject variance.

Both language groups performed above chance in all sentence conditions (one-sample t-tests: all $t > 6.79$, all $p < .001$). All the children attained a minimum of 62.5% of correct responses in the active sentences.

A first LMM was run, in order to investigate whether the performance of children differentiated in the sentence conditions. The dependent variable of the model was the number of correct responses and fixed factors were group (i.e., Bili-Mono), and condition (ACT PF/ACT MF/PAS PF/PAS MF). The model included two random intercepts: participants and items, and a random slope: the factor condition which is allowed to vary for each participant. A LMM was computed (see Appendix E.1 for complete model output), and a treatment contrast was set for language group and condition, in which subsets of the data, i.e., monolinguals and PAS PF, were coded to be in the intercept. Note that in the fixed effect column of the LMM tables, the subset of data to which the simple effect refers to was written in parentheses. This will be the case for all further analyses. Table 32 illustrates parameters of the LMM: estimates, SE, *z*-values and *p*-values are given as the fixed factors.

Table 32

Parameters of LMM for correct responses with monolinguals and PAS PF in the intercept

Fixed effect	**Estimate**	**SE**	***z*-value**	***p*-value**
Intercept (Mono, PAS PF)	2.364	0.273	8.658	<.001***
PAS PF (Bili)	-0.666	0.353	-1.888	=.059±
PAS MF (Mono)	-0.584	0.328	-1.783	=.075±
ACT PF (Mono)	0.257	0.394	0.653	=.514
ACT MF (Mono)	-0.603	0.330	-1.825	=.068±
Bili x PAS MF	-0.066	0.425	-0.154	=.878
Bili x ACT PF	0.646	0.531	1.217	=.224
Bili x ACT MF	0.661	0.437	1.511	=.131

± $p < .1$, *** $p < .001$

Estimates indicated the difference scores of the logit-transformed proportion of correct responses (zero = 1:1 = 50%). The number of correct responses for PAS PF in monolinguals was above chance ($p < .001$), and bilinguals did not give significantly less correct responses than monolinguals ($p = .059$). Within the monolingual group, there were no significant differences between the sentence conditions (all $p > .068$).

To evaluate whether the number of correct responses in the bilingual group was different between conditions and whether there were any differences between the groups' performance in the PAS MF sentence condition, a further LMM was run with different contrast coding (bilinguals and PAS MF coded to be in the intercept). All other model specifications were the same as in the previous model. Table 33 reports the results of the model (for complete output, see Appendix E.2).

Table 33

Parameters of LMM for correct responses with bilinguals and PAS MF in the intercept

Fixed effect	**Estimate**	**SE**	***z*-value**	***p*-value**
Intercept (Bili, PAS MF)	1.048	0.194	5.414	<.001***
PAS MF (Mono)	0.729	0.291	2.508	=.012*
PAS PF (Bili)	0.658	0.272	2.414	=.016*
ACT PF (Bili)	1.548	0.321	4.824	<.001***
ACT MF (Bili)	0.706	0.258	2.734	=.006**
Mono x PAS PF	-0.058	0.428	-0.135	=.893
Mono x ACT PF	-0.714	0.461	-1.548	=.122
Mono x ACT MF	-0.725	0.374	-1.938	=.053±

± $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$

The number of correct responses for PAS MF in bilinguals was above chance level ($p < .001$), and monolinguals gave significantly more correct responses than bilinguals in this condition ($p = .012$). Within the bilingual group and with respect to the PAS MF condition, a higher number of correct responses were given for the three other sentence conditions (all $p < .016$).

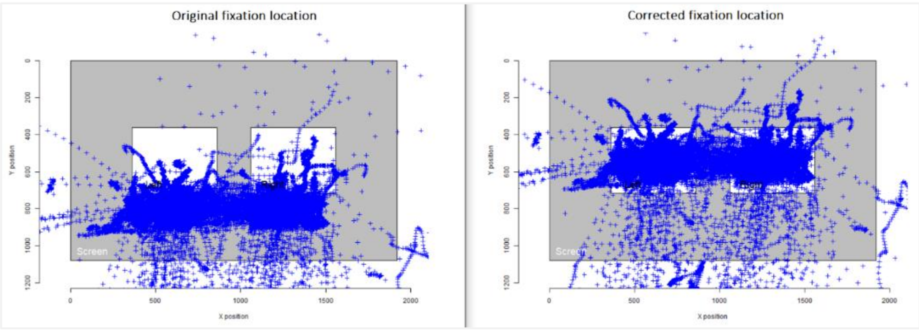
In sum, these results demonstrate that all the children performed above chance in the comprehension of passive sentences. The performance of monolingual children did not differ between sentence conditions, whereas bilinguals performed with lower scores on PAS MF than on all other sentence conditions. The performance of monolinguals and bilinguals is comparable in the PAS PF condition, while monolinguals outperformed bilinguals in the PAS MF condition.

6.2.3.2 Pre-processing of eye-gaze data

All blinks and data points with looks outside the screen were removed. The remaining data for the whole groups contained an average of 87.4% (range: 56.7%-96.6%) looks to the AoIs (50% to the target and 37.4% to the distractor). As in Study 1, the median offset for each participant's X and Y gaze positions was measured, by using the central fixation point at the beginning of each trial (averaged for the whole experiment). This value was then applied to calculate a corrected gaze position. Subsequently, each participant's original and corrected gaze positions were plotted for the evaluation of the results (for an example see Figure 27).

Figure 27

Example of original and corrected recorded eye fixation location of one 4-year-old child



Note. The x-axes represent the X coordinates of eye gaze positions and the y-axes the Y coordinates on the screen recorded during the eye-tracking experiment. Original fixation location is represented on the left image and corrected fixation location on the right.

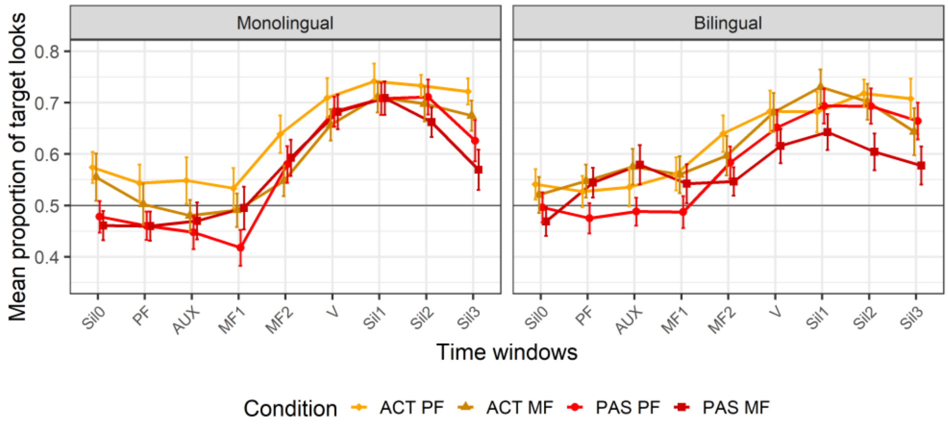
As earlier, a cutoff of 10% was used: if the number of data points falling into one of the two AoIs increased by at least 10% in the corrected gaze position, compared to the original gaze position, then the corrected gaze position was used. This was the case for two of the children. The original fixation location in one child into the AoIs was 80.1% and increased to 94% after correction of fixation, while in the second case, original fixation was 42.1% and increased to 89.1% after correction of fixation (as shown in Figure 27 above). In 19 children the increase in gazes to the AoI amounted to less than 1.6%, in two others, the correction led to an increase of 2.5% and 3.4%. For these children, the original gaze positions were used according to the set criterion.

6.2.3.3 Eye-gaze data

Mean proportion of looks to the target picture was calculated by dividing the looking proportion to the target by the sum of looking proportion to the target and to the distractor. Eye-gaze data were (proportion of looks to the target picture) first averaged for each time window (i.e., PF, AUX, MF1, MF2, V, Sil1, Sil2, Sil3) in every trial, then across all trials of each condition (ACT PF, ACT MF, PAS PF and PAS MF) and afterwards across participants within each group (monolinguals and bilinguals). Sil0 indicated the time of 750 ms of silence until the onset of PF, time window PF indicated the mean length between PF onset and PF offset, time window AUX the mean length between AUX onset and AUX offset and so on. Figure 28 plots the mean proportion of looks to the target (y-axis) for conditions during the nine time windows (x-axis).

Figure 28

Proportion of looks to the target for ACT PF, ACT MF, PAS PF and PAS MF conditions



Note. The left panel shows the proportion of looks to the target in the monolingual group and the right panel in the bilingual group. The gray horizontal line at 0.5 indicates chance level. Error bars represent ±1 SE of between-subject variance.

LMMs with participants and items as random effects were fitted. Random slopes led the model to non-convergence, so they had to be dropped. Fixed effects included language group, condition and the interaction of the two. Looks to the target picture were modelled for each time window separately. To compare the proportion of looks to the target across conditions and groups, a treatment contrast coding was applied. First, monolinguals and PAS PF were set to the intercept. The intercept represents the grand mean across language groups and conditions. The values of the intercept are not of interest, since they represent the predicted value of the dependent variable when all independent variables are zero, averaged across all fixed effects. Table 34 provides the parameters for the LMMs (see Appendix E.3-E.10 for complete model outputs).

Table 34

Parameters for the LMMs for eye-gaze data with monolinguals and PAS PF in the intercept

Time window	**Fixed effect**	**Estimate**	**SE**	**df**	***t*-value**	***p*-value**
PF	Intercept (Mono, PAS PF)	0.457	0.033	647.79	13.393	<.001***
	PAS PF (Bili)	0.017	0.047	1443	0.357	=.721
	PAS MF (Mono)	0.002	0.047	1423	0.051	=.960
	ACT PF (Mono)	0.071	0.047	669.27	1.508	=.132
	ACT MF (Mono)	0.046	0.047	678.95	0.979	=.328
	Bili x PAS MF	0.060	0.067	1460.6	0.900	=.368

Time window	Fixed effect	Estimate	SE	df	*t*-value	*p*-value
	Bili x ACT PF	-0.019	0.067	1440.3	-0.278	=.781
	Bili x ACT MF	0.021	0.067	1444.4	0.749	=.749
AUX	Intercept (Mono, PAS PF)	0.514	0.035	1414	12.716	<.001***
	PAS PF (Bili)	0.036	0.050	1414	0.722	=.470
	PAS MF (Mono)	0.034	0.051	1414	0.671	=.502
	ACT PF (Mono)	0.087	0.050	1414	1.740	=.082‡
	ACT MF (Mono)	0.031	0.051	1414	0.599	=.550
	Bili x PAS MF	0.063	0.072	1414	0.871	=.384
	Bili x ACT PF	-0.044	0.072	1414	-0.609	=.543
	Bili x ACT MF	0.053	0.072	1414	0.729	=.466
MF1	Intercept (Mono, PAS PF)	0.421	0.034	1484	12.222	<.001***
	PAS PF (Bili)	0.058	0.049	1484	1.185	=.236
	PAS MF (Mono)	0.082	0.049	1484	1.698	=.090‡
	ACT PF (Mono)	0.110	0.049	1484	2.263	=.024*
	ACT MF (Mono)	0.063	0.049	1484	1.296	=.195
	Bili x PAS MF	-0.018	0.069	1484	-0.262	=.794
	Bili x ACT PF	-0.033	0.069	1484	-0.474	=.635
	Bili x ACT MF	0.020	0.069	1484	0.296	=.767
MF2	Intercept (Mono, PAS PF)	0.587	0.031	6673.1	18.682	<.001***
	PAS PF (Bili)	-0.001	0.044	1489	-0.032	=.974
	PAS MF (Mono)	0.006	0.042	1478	0.133	=.894
	ACT PF (Mono)	0.051	0.043	66.82	1.158	=.247
	ACT MF (Mono)	-0.037	0.045	678.4	-0.836	=.403
	Bili x PAS MF	-0.043	0.063	1516	-0.694	=.488
	Bili x ACT PF	-0.002	0.063	1496	-0.039	=.969
	Bili x ACT MF	0.053	0.063	1498	0.841	=.400
V	Intercept (Mono, PAS PF)	0.687	0.035	273.3	19.666	<.001***
	PAS PF (Bili)	-0.035	0.049	309.7	-0.729	=.467

Time window	**Fixed effect**	**Estimate**	**SE**	**df**	***t*-value**	***p*-value**
	PAS MF (Mono)	-0.003	0.044	1421	-0.077	=.939
	ACT PF (Mono)	0.025	0.045	548.8	0.557	=.578
	ACT MF (Mono)	-0.032	0.045	553.2	-0.711	=.478
	Bili x PAS MF	-0.031	0.062	1451	-0.484	=.629
	Bili x ACT PF	0.008	0.062	1433	0.126	=.900
	Bili x ACT MF	0.071	0.063	1437	1.124	=.261
Sil1	Intercept (Mono, PAS PF)	0.710	0.034	250.0	20.819	<.001***
	PAS PF (Bili)	-0.019	0.048	246.6	-0.387	=.699
	PAS MF (Mono)	0.000	0.042	1476	0.008	=.994
	ACT PF (Mono)	0.033	0.042	1475	0.796	=.426
	ACT MF (Mono)	0.002	0.042	1474	0.051	=.959
	Bili x PAS MF	-0.051	0.059	1475	-0.829	=.385
	Bili x ACT PF	-0.042	0.059	1476	-0.704	=.482
	Bili x ACT MF	0.043	0.059	1475	0.731	=.465
Sil2	Intercept (Mono, PAS PF)	0.708	0.033	299.6	21.582	<.001***
	PAS PF (Bili)	-0.015	0.046	342.6	-0.250	=.803
	PAS MF (Mono)	-0.049	0.043	1403	-1.155	=.248
	ACT PF (Mono)	0.025	0.042	637.9	0.601	=.548
	ACT MF (Mono)	-0.007	0.043	649.1	-0.171	=.864
	Bili x PAS MF	0.046	0.060	1431	-0.768	=.443
	Bili x ACT PF	-0.011	0.060	1408	-0.177	=.860
	Bili x ACT MF	0.020	0.060	1412	-0.335	=.738
Sil3	Intercept (Mono, PAS PF)	0.628	0.034	443.02	18.581	<.001***
	PAS PF (Bili)	0.025	0.047	432.10	0.535	=.593
	PAS MF (Mono)	-0.040	0.047	1320	-0.857	=.392
	ACT PF (Mono)	0.094	0.046	1313	2.052	=.040
	ACT MF (Mono)	0.048	0.046	1318	1.044	=.297
	Bili x PAS MF	-0.012	0.065	1321	-0.182	=.856
	Bili x ACT PF	-0.050	0.065	1315	-0.764	=.445

Time window	Fixed effect	Estimate	SE	df	*t*-value	*p*-value
	Bili x ACT MF	-0.052	0.062	1317	-0.816	=.414

‡ *p* <.1, * *p* <.05, *** *p* < .001

On comparing the PAS PF condition in monolingual and bilingual children during sentence processing, no statistically significant differences between the two groups were observed. The only difference within the monolingual group occurred during time window MP1: the proportion of looks to the target was larger in ACT PF compared to PAS PF.

Finally, LMMs with different contrast coding were run (bilinguals and PAS MF coded to be in the intercept), to evaluate whether the proportion of target looking in the bilingual group differed between conditions and to the monolingual group compared to PAS PF. Further model specifications were the same as in previous models. Table 35 reports these parameters (see Appendix E.11-E.18 for complete model outputs).

Table 35

Parameters for the LMMs for eye-gaze data with bilinguals and PAS MF in the intercept

Time window	Fixed effect	Estimate	SE	df	*t*-value	*p*-value
PF	Intercept (Bili, PAS MF)	0.537	0.033	649.6	15.919	<.001***
	PAS MF (Mono)	-0.077	0.048	1443	-1.607	=.108
	PAS PF (Bili)	-0.063	0.047	1465	-1.321	=.187
	ACT PF (Bili)	-0.010	0.048	644.5	-0.215	=.830
	ACT MF (Bili)	0.005	0.048	642.3	0.108	=.914
	Mono x PAS PF	0.060	0.067	1461	0.900	=.368
	Mono x ACT PF	0.079	0.068	1440	1.168	=.243
	Mono x ACT MF	0.039	0.068	1444	0.573	=.566
AUX	Intercept (Bili, PAS MF)	0.582	0.036	1414	16.169	<.001***
	PAS MF (Mono)	-0.099	0.051	1414	-1.925	=.055‡
	PAS PF (Bili)	-0.097	0.051	1414	-1.907	=.057‡
	ACT PF (Bili)	-0.053	0.051	1414	-1.029	=.304
	ACT MF (Bili)	-0.013	0.051	1414	-0.259	=.796
	Mono x PAS PF	0.063	0.072	1414	0.871	=.384

Time window	**Fixed effect**	**Estimate**	**SE**	**df**	***t*-value**	***p*-value**
	Mono x ACT PF	0.106	0.073	1414	1.464	=.143
	Mono x ACT MF	0.010	0.073	1414	0.134	=.893
MF1	Intercept (Bili, PAS MF)	0.543	0.034	1484	15.903	<.001***
	PAS MF (Mono)	-0.040	0.049	1484	-0.815	=.415
	PAS PF (Bili)	-0.065	0.049	1484	-1.340	=.181
	ACT PF (Bili)	0.012	0.049	1484	0.254	=.800
	ACT MF (Bili)	0.019	0.049	1484	0.383	=.702
	Mono x PAS PF	-0.018	0.069	1484	-0.262	=.794
	Mono x ACT PF	0.015	0.069	1484	0.213	=.832
	Mono x ACT MF	-0.038	0.069	1484	-0.557	=.577
MF2	Intercept (Bili, PAS MF)	0.548	0.032	648.07	17.303	<.001***
	PAS MF (Mono)	0.045	0.044	1498	1.011	=.312
	PAS PF (Bili)	0.038	0.044	1518	0.848	=.396
	ACT PF (Bili)	0.086	0.045	656.02	1.922	=.055‡
	ACT MF (Bili)	0.053	0.045	662.05	1.179	=.239
	Mono x PAS PF	-0.043	0.063	1516	-0.694	=.488
	Mono x ACT PF	-0.041	0.063	1496	-0.652	=.514
	Mono x ACT MF	-0.096	0.063	1497	-1.527	=.127
V	Intercept (Bili, PAS MF)	0.618	0.035	273.3	17.567	<.001***
	PAS MF (Mono)	0.066	0.049	309.7	1.350	=.178
	PAS PF (Bili)	0.034	0.044	1455	0.757	=.449
	ACT PF (Bili)	0.066	0.046	545.8	1.456	=.146
	ACT MF (Bili)	0.072	0.046	555.3	1.565	=.118
	Mono x PAS PF	-0.030	0.062	1451	-0.484	=.629
	Mono x ACT PF	-0.038	0.063	1432	-0.609	=.543
	Mono x ACT MF	-0.101	0.063	1435	-1.604	=.109
Sil1	Intercept (Bili, PAS MF)	0.641	0.034	245.0	18.897	<.001***
	PAS MF (Mono)	0.070	0.048	247.0	1.451	=.148

Time window	**Fixed effect**	**Estimate**	**SE**	**df**	***t*-value**	***p*-value**
	PAS PF (Bili)	0.051	0.042	1474	1.227	=.220
	ACT PF (Bili)	0.042	0.042	1476	1.011	=.312
	ACT MF (Bili)	0.096	0.042	1475	2.294	=.022*
	Mono x PAS PF	-0.051	0.059	1475	-0.869	=.385
	Mono x ACT PF	-0.009	0.059	1476	-0.160	=.873
	Mono x ACT MF	-0.095	0.059	1475	-1.592	=.111
Sil2	Intercept (Bili, PAS MF)	0.601	0.033	293.5	18.324	<.001***
	PAS MF (Mono)	0.058	0.046	350.6	1.244	=.214
	PAS PF (Bili)	0.095	0.043	1430	2.258	=.024*
	ACT PF (Bili)	0.110	0.043	620.9	2.578	=.010*
	ACT MF (Bili)	0.108	0.043	607.0	2.538	=.011*
	Mono x PAS PF	-0.046	0.060	1431	-0.768	=.443
	Mono x ACT PF	-0.036	0.060	1411	-0.592	=.554
	Mono x ACT MF	-0.066	0.060	1413	-1.100	=.271
Sil3	Intercept (Bili, PAS MF)	0.602	0.034	426.79	17.833	<.001***
	PAS MF (Mono)	-0.013	0.048	445.37	-0.280	=.780
	PAS PF (Bili)	0.052	0.046	1322	1.134	=.257
	ACT PF (Bili)	0.960	0.047	1316	2.063	=.039*
	ACT MF (Bili)	0.047	0.046	1316	1.030	=.303
	Mono x PAS PF	-0.012	0.065	1321	-0.182	=.856
	Mono x ACT PF	0.038	0.066	1316	0.576	=.565
	Mono x ACT MF	0.041	0.065	1317	0.626	=.531

‡ p <.1, * p <.05, *** p < .001

Differences between the two groups were not detected here either: during the processing of PAS MF sentences, the two groups displayed the same looking behavior in all time windows. Differences within the bilingual group were observed after sentence offset: in Sil1 the proportion of looks to the target was larger in ACT MF than in PAS MF, in Sil2, the bilinguals looked more to the target during PAS PF, ACT PF and ACT MF than in PAS MF, and in Sil3 they showed larger proportion of looks to the target during ACT PF than in PAS MF.

In summary, the eye-tracking data did not show any differences between the two language groups. After the voice cue monolingual children looked more to the target during active sentences than in passive sentences with the subject in the prefield. However, this difference is only present during the time window MF1, i.e., the adverb. After MF1, the processing of all sentences is comparable within the monolingual group. Within the bilingual group, the proportion of looks to the target during the process of passives with the subject in the middlefield was smaller than during the process of actives with the subject in the middlefield in Sil1 and Sil2. During Sil2, the main difference was between the two passive conditions: bilingual children looked less to the target during passive sentences with the subject in the middlefield than in the other sentence conditions. In Sil3, this difference is still only present between PAS MF and ACT PF.

As in Study 1, further analyses were also run to investigate whether participants' gaze data predicted their responses or showed any inconsistency between offline and online results. Before proceeding with the next LMM, individual data are presented. The numbers of participants to have given a specific number of correct responses in each sentence condition are reported in the following Table.

Table 36

Individual data for correct responses in the four passive sentence conditions

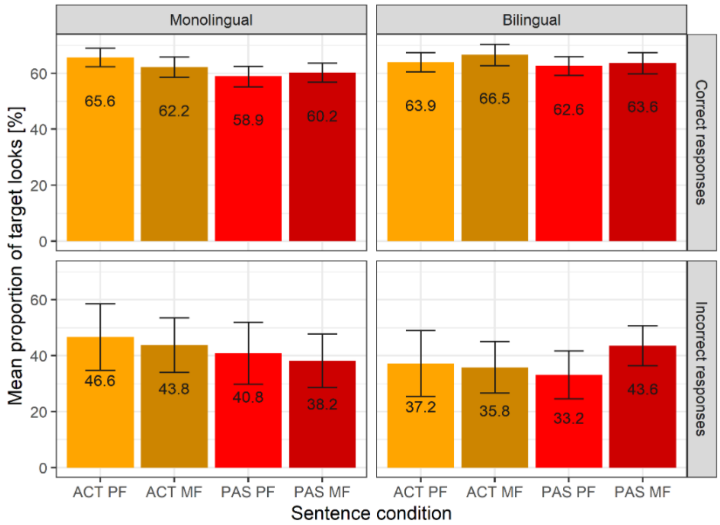
Condition	**Group**	**8/8** correct (100%)	**7/8** correct (87.5%)	**6/8** correct (75%)	**5/8** correct (62.5%)	**4/8** correct (50%)
ACT PF	Monolingual	13	9	2	–	2
	Bilingual	13	8	3	1	1
ACT MF	Monolingual	7	10	6	2	1
	Bilingual	4	14	7	–	1
PAS PF	Monolingual	13	8	2	3	–
	Bilingual	8	6	8	2	2
PAS MF	Monolingual	8	7	7	4	–
	Bilingual	5	2	6	9	4

Note. The total number of correct responses in the ACT PF condition for all 26 monolinguals and all 26 bilinguals is 187 out of 208 (89.9%). In the ACT MF condition, the number of correct responses by all 26 monolinguals and all 26 bilinguals is 176 out of 208 (84.6%). In the PAS PF condition, the total number of correct responses by monolinguals is 187 out of 208 (89.9%), as in the ACT PF condition, and by bilinguals is 172 out of 208 (82.7%). In the PAS MF condition, the number of correct responses by monolinguals is 175 out of 208 (84.1%) and by bilinguals is 151 out of 208 (72.6%).

The number of individual children's correct responses in active and passive sentences was at least 50%. Figure 29 below presents the percentage of mean proportion of target looks in the two groups, separated by correctness of behavioral responses.

Figure 29

Proportion of target looking for ACT PF, ACT MF, PAS PF and PAS MF, broken down by group and responses



Note. The mean proportion of looking to the target are shown in the y-axis, and sentence conditions are reported on the x-axis. Error bars represent ±1 SE of between-subject variance.

The mean proportion of looking to the target was used as the dependent variable across time windows in the next LMM (Sil0 was excluded from the model since it included 750 ms before sentence onset). Trials containing correct responses were set as the reference level (i.e., intercept), using the treatment contrast. Fixed effects were: language group, condition and accuracy. Three random components were set up: participants and items as random intercepts and the effect of condition for each participant as a random slope. Sentence conditions, as well as the language group, were coded with a repeated contrast. The contrast ACT MF-ACT PF with a positive estimate means that a higher number of looks to the target took place for ACT MF than for ACT PF, since ACT PF was subtracted from ACT MF. The same contrast with a negative estimate signals more target looks for ACT PF. In the contrast PAS PF-ACT MF, ACT MF was subtracted from PAS PF, i.e., positive estimates indicate more looks to the target for PAS PF, and negative estimates indicate more looks to the target for ACT MF. In the contrast Bili-Mono, monolinguals were subtracted from bilinguals. More looks to the target for bilinguals are indicated by a positive estimate and for monolinguals by a negative estimate. Parameters of the models are reported in Table 37 (see Appendix E.19 for complete model outputs). Only significant effects are discussed.

Table 37

Parameters of the LMM for passive eye-gaze data with correct responses in the intercept

Fixed effect	**Estimate**	**SE**	**df**	***z*-value**	***p*-value**
Intercept (Corr. resp, grand mean)	0.646	0.008	75.12	76.077	<.001***
ACT MF-ACT PF	-0.002	0.019	54.69	-0.079	=.938
PAS PF-ACT MF	-0.036	0.022	71.51	-1.592	=.012
PAS MF-PAS PF	0.019	0.016	58.32	1.170	=.247
Bili-Mono	-0.028	0.014	54.44	1.982	=.053‡
Incorr. resp	-0.273	0.012	8893	-21.849	<.001***
ACT MF-ACT PF x Bili-Mono	0.066	0.039	54.97	1.690	=.097‡
PAS PF-PAS MF x Bili-Mono	-0.012	0.041	54.58	-0.305	=.762
PAS MF-PAS PF x Bili-Mono	-0.014	0.033	58.45	-0.415	=.680
ACT MF-ACT PF x incorr. resp	-0.054	0.037	6757	-1.452	=.146
PAS PF-ACT MF x incorr. resp	0.025	0.035	6833	0.720	=.471
PAS MF-PAS PF x incorr. resp	-0.021	0.034	6221	-0.619	=.536
Bili-Mono x incorr. resp	-0.062	0.025	9260	-2.513	=.012*
ACT MF-ACT PF x Bii-Mono x incorr. resp	-0.074	0.074	7365	-0.994	=.320
PAS PF-ACT MF x Bii-Mono x incorr. resp	0.008	0.069	7123	0.112	=.911
PAS MF-PAS PF x Bii-Mono x incorr. resp	0.119	0.066	6502	1.788	=.074‡

‡ p <.01, * p <.05, *** p < .001

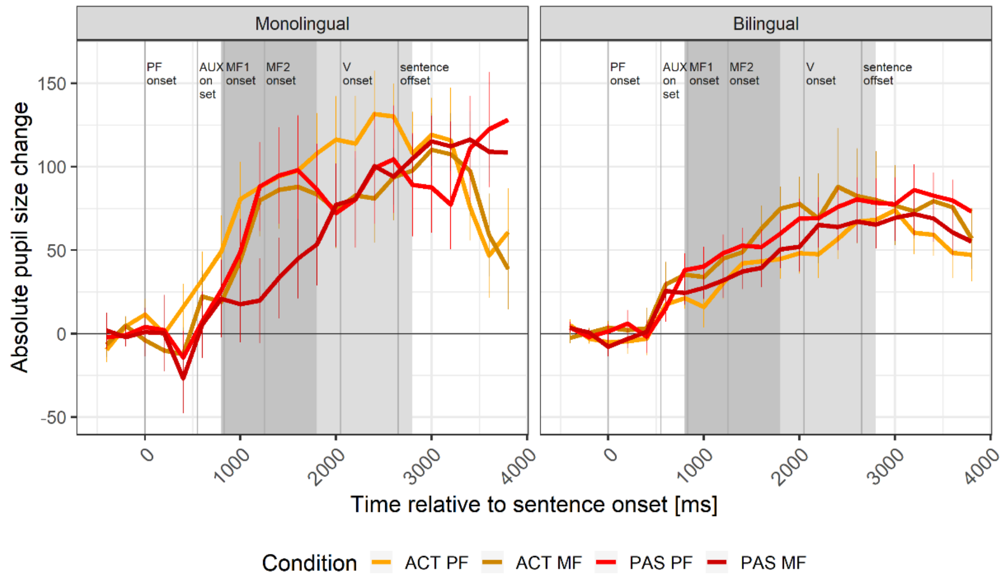
Across conditions children looked less to target when giving incorrect responses, compared to correct responses (p < .001). The interaction Bili-Mono x incorr. resp gives evidence that the number of looks to the target was higher in correct trials compared to incorrect trials and in the monolingual compared to the bilingual group. This is reflected in Figure 29 by the fact that monolingual children looked more to the target in incorrect trials (42.4%) than the bilingual children did (37.5%). The purpose of this analysis was to investigate the relation between offline and online measures. Thus, it can be demonstrated that eye gazes predicted children's responses.

6.2.3.4 Pupillometry

Pupil size data across time was provided by the eye-tracking system for all data points containing gaze positions (across all non-blinks). The measured pupil size area consists of data in arbitrary units[11] (of between 100 and 10000) and recorded in scaled camera image pixels, since pupils are not calibrated and the units of pupil measurement vary with each calibration and thus for each participant. In each trial, a baseline pupil dilation was calculated at an interval of 500 ms, just before sentence onset, i.e., during silence (Sil0). The proportion of pupil dilation change with respect to the baseline was calculated by subtracting the baseline value from the measured pupil size within each trial. Trials with more than 25% missing data were removed. This affected 19.2% of the data. Eight children out of 62 had less than 25% missing pupil data in all trials. On average, a mean of 6.9 trials per participant had to be removed (range: 1-26 trials). Missing points were not interpolated. Figure 30 shows the time course of the averaged pupil dilation responses, with average sentence constituent borders marked by vertical lines.

Figure 30

Pupillometry data for passive conditions broken down by group



Note. The dark gray areas in the graphs indicate time window 1 (800-1800 ms); the light gray areas indicate time window 2 (1800-2800 ms). The y-axis represents average pupil dilation with respect to the baseline 0 in arbitrary units; the x-axis shows the 3800 ms time relative to sentence onset. Vertical lines indicate the average onsets of the sentence constituents (PF marks the sentence onset at 0 ms, the onset of AUX had a mean of 550 ms, the onset of MF1 had a mean of 800 ms; the onset of MF2 had a mean of 1350 ms and the onset of the lexical verb 2150 ms). Error bars represent ±1 SE of between-subject variance.

[11] Pupil size areas, averaged across trials and then across participants, showed that the monolingual group had larger pupil size areas (1539 in arbitrary units) than the bilingual group (850 in arbitrary units), and the Mann-Whitney-U-Test showed a significant difference ($W = 0$, $p < .001$).

As in Study 1, monolingual and bilingual children were tested under different light conditions, which is why group differences are not taken into account.

Considering that the pupil reaches its maximum deflection between one and two seconds after the evoking event (Just & Carpenter, 1993) and that children's pupils may need more time than in adults to reach their maximum size, two time windows were considered after the offset of the voice cue. Pupil size changes were analyzed after AUX offset (i.e., the voice cue to passive), approximately 800 ms after sentence onset, in two time windows: the first second after AUX offset (time window 1 between 800 ms and 1800 ms) and the second one (time window 2 between 1800 ms and 2800 ms). The fixed effect is condition. A repeated contrast was used for condition: ACT MF-ACT PF with positive estimates means that there was larger pupil dilation for ACT MF and with negative estimates for ACT PF; the contrast PAS PF-ACT MF with positive estimates indicates larger pupil dilation for PAS PF and negative estimates for ACT MF; the last contrast for condition PAS MF-PAS PF with positive estimates indicates larger pupil dilation for PAS MF and negative estimates for PAS PF. First, the monolinguals' data were analyzed. Random effects were participants and items (see Appendix E.20-E.21 for complete model outputs). The Table 38 below reports the parameters.

Table 38

Parameters for the LMMs for pupillometry with 4-year-old monolinguals

Time window	**Fixed effect**	**Estimate**	**SE**	**df**	***t*-value**	***p*-value**
1	Intercept (grand mean)	76.588	16.74	46.13	4.576	<.001***
1	ACT MF-ACT PF	-18.769	1.250	323997	-15.02	<.001***
1	PAS PF-ACT MF	19.371	18.05	61.758	1.073	=.287
1	PAS MF-PAS PF	-49.263	1.267	324018	-38.87	<.001***
2	Intercept (grand mean)	101.76	18.30	43.07	5.56	<.001***
2	ACT MF-ACT PF	-32.35	1.353	324480	-23.91	<.001***
2	PAS PF-ACT MF	3.842	18.67	61.687	0.206	=.838
2	PAS MF-PAS PF	2.979	1.365	324504	2.183	<.001***

*** $p < .001$

During both time windows, the pupil dilation in monolingual children was larger for the ACT PF condition than for the ACT MF condition ($p < .001$). The contrast PAS MF-PAS PF showed a larger pupil dilation in the PAS PF condition compared to the PAS MF condition in time window 1 ($p < .001$) and a reverse pattern in time window 2, namely pupil dilation for PAS MF, which was larger than for PAS PF ($p < .001$).

In the succeeding analysis, the bilingual children's data were analyzed (see Appendix E.22-E.23 for model output). Sentence conditions were coded by a repeated contrast, and random effects were participants and items, as in the previous pupillometry analysis.

Table 39

Parameters for the LMMs for pupillometry in 4-year-old bilinguals

Time window	**Fixed effect**	**Estimate**	**SE**	**df**	***t*-value**	***p*-value**
1	Intercept grand mean)	44.16	8.434	42.42	5.237	<.001***
1	ACT MF-ACT PF	14.95	0.670	334200	22.331	<.001***
1	PAS PF-ACT MF	1.657	8.445	6.179	0.196	=.845
1	PAS MF-PAS PF	-13.02	0.647	334200	-20.139	<.001***
2	Intercept (grand mean)	68.77	14.07	43.49	4.887	<.001***
2	ACT MF-ACT PF	23.66	95.01	334700	24.898	<.001***
2	PAS PF-ACT MF	-5.212	14.42	61.17	-0.361	=.719
2	PAS MF-PAS PF	-11.57	0.923	334700	-12.540	<.001***

* p <.05, *** p < .001

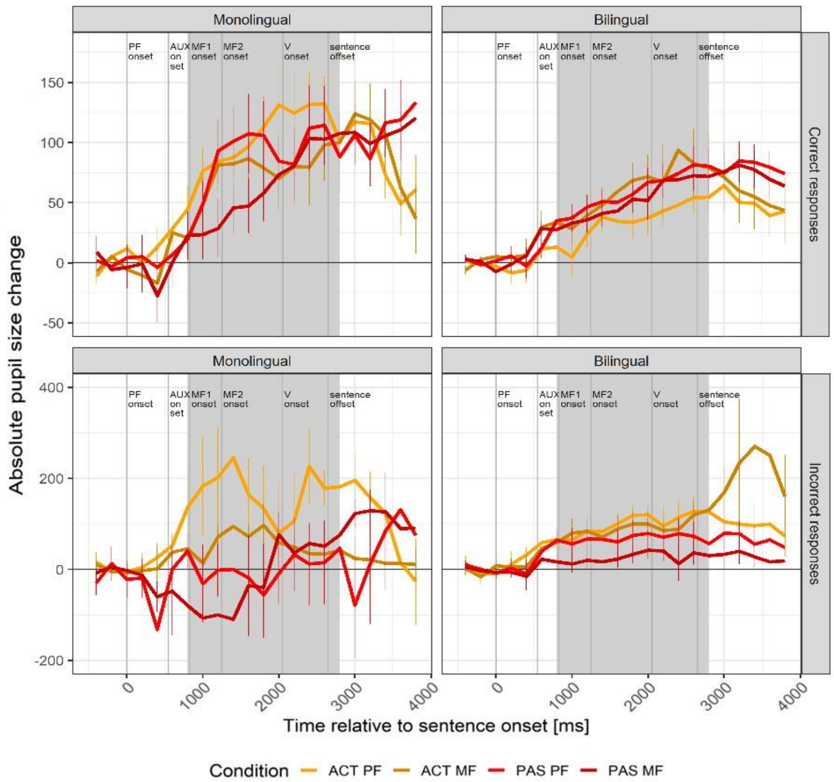
In time windows 1 and 2, pupil dilation in the bilingual group was larger in the ACT MF condition than in the ACT PF condition and again larger in PAS PF than in PAS MF (all $p < .001$). For the contrast ACT MF-ACT PF, the bilingual group displayed a reverse pattern, in comparison to the monolingual group: for both time windows, pupil dilation was smaller in the ACT PF condition than in the ACT MF condition ($p < .001$). Interaction effects reflected those of the previous analysis.

To summarize, throughout the processing of passive sentences, the monolingual group showed larger pupil dilation for PAS PF than PAS MF sentences immediately after the offset of AUX (time window 1) and larger pupil dilation for PAS MF than PAS PF during hearing the *by*-phrase (time window 2). In contrast, bilingual children displayed pupil dilation during the processing of PAS PF sentences that was larger than that of PAS MF sentences in both time windows.

Analogously to Study 1, it was examined whether there could be a connection between the online measurement of pupillometry and accuracy data. It might be the case that the very different pupillary response patterns (between monolinguals and bilinguals, or within conditions) vary less when accounting for accuracy. Thus, to investigate whether pupil data was related to participants' responses, a further analysis was run, this time adding the factor of accuracy to the model. Figure 31 presents the time course of the average pupil dilation for correct and incorrect responses for both groups.

Figure 31

Pupil dilation in different passive conditions broken down by group and responses



Note. The gray area in the graphs indicate the only time window between 800-2800 ms. The y-axis represents the average pupil dilation with respect to the baseline 0 in arbitrary units. The x-axis illustrates 3800 ms time relative to sentence onset. Please note the different scales on the y-axis for correct responses (upper panels) and incorrect responses (lower panels) to represent the difference in pupil size changes between conditions more clearly. Error bars represent ±1 SE of between-subject variance.

The pupil data included in the following analysis were those in time window 1 and 2 (between 800 ms and 2800 ms). Both time windows were combined into one window for this analysis. Repeated contrast was used for condition, whereas correct responses were coded with treatment contrast and set as the reference level (i.e., intercept). Two random components were included: participants and items as random intercepts (see Table 40 and Appendix E.24 for complete model output). First, the monolinguals' data were analyzed.

Table 40

Parameters for the LMMs for the passive pupil data with correct responses for monolinguals

Fixed effect	**Estimate**	**SE**	**df**	***t*-value**	***p*-value**
Intercept (Corr. resp, grand mean)	95.812	17.28	44.054	5.545	<.001***
ACT MF-ACT PF (corr. resp)	-22.51	0.992	648557	-22.69	<.001***
PAS PF-ACT MF (corr. resp)	25.78	17.93	61.631	1.438	=.156
PAS MF-PAS PF (corr. resp)	-20.51	1.001	648562	-20.48	<.001***
Incorr. resp (corr. resp)	-46.05	1.120	648507	-41.12	<.001***
ACT MF-ACT PF x Incorr. resp	-39.87	3.239	648353	-12.31	<.001***
PAS PF-ACT MF x Incorr. resp	-133.33	3.134	648326	-42.55	<.001***
PAS MF-PAS PF x Incorr. resp	12.847	3.078	648543	4.174	<.001***

*** $p < .001$

In their correct responses, the monolinguals' pupil dilation was larger in the ACT PF condition than in the ACT MF condition, and, again, larger in PAS PF than in PAS MF (all $p < .001$). The monolinguals' pupil dilation was larger in correct responses than in incorrect ones across conditions ($p < .001$). Interaction effects in the monolingual group revealed that the pupil dilation difference between ACT MF-ACT PF and between PAS PF-ACT MF was smaller in incorrect responses than in correct responses ($p < .001$), whereas the pupil dilation difference between PAS MF-PAS PF was larger in incorrect responses than in correct responses ($p < .001$).

The next analysis contained bilinguals and correct responses in the intercept. The factor condition was coded with repeated contrast. Random effects were coded as in the previous model (see Appendix E.25). Some effects of this model are not discussed, since they reflected the results of the previous model.

Table 41

Parameters for the LMMs for the passive pupil data with the correct responses for bilinguals

Fixed effect	**Estimate**	**SE**	**df**	***t*-value**	***p*-value**
Intercept (Corr. resp, grand mean)	54.69	10.99	41.01	4.979	<.001***
ACT MF-ACT PF (corr. resp)	22.44	0.631	668900	35.583	<.001***

Fixed effect	**Estimate**	**SE**	**df**	***t*-value**	***p*-value**
PAS PF-ACT MF (corr. resp)	2.737	10.66	61.66	0.257	=.798
PAS MF-PAS PF (corr. resp)	-7.757	6.445	668900	-12.034	<.001***
Incorr. resp (corr. resp)	19.43	0.598	668900	32.499	<.001***
ACT MF-ACT PF x Incorr. resp	-41.02	1.821	668900	-22.52	<.001***
PAS PF-ACT MF x Incorr. resp	-31.14	1.657	668900	-18.79	<.001***
PAS MF-PAS PF x Incorr. resp	-16.66	1.508	668900	-11.05	<.001***

* p <.05, ** p <.01, *** p < .001

In their correct responses, the bilinguals' pupil dilation proved to be larger in the ACT MF condition than in the ACT PF condition, and again, larger in PAS PF than in PAS MF (all $p < .001$). The bilinguals' pupil dilation was larger in incorrect responses than in correct responses across conditions ($p < .001$). Pupil dilation difference in the bilingual group between ACT MF-ACT PF and between PAS PF-ACT MF and PAS MF-PAS PF was smaller in incorrect responses than in correct responses ($p < .001$).

In sum, when giving correct responses bilingual children revealed larger pupil dilation during the processing of ACT MF than that of ACT PF, as well as larger during the processing of PAS PF than that of PAS MF sentences. When giving correct responses, the monolingual children displayed larger pupil dilation during the processing of ACT PF than in that of ACT MF, and also larger during the processing of PAS PF than PAS MF sentences. The monolinguals' pupil dilation was larger in correct responses than in incorrect responses across conditions ($p < .001$), whereas the bilinguals' pupil dilation was larger in incorrect responses than in correct responses across conditions ($p < .001$).

6.2.4 Executive function tasks

<u>Flanker task</u>

As in Study 1, both the practice items and the test items with missing responses or response latency of longer than 4000 ms and shorter than 200 ms were removed from the Flanker task analysis. This affected 1.2% of the data. Furthermore, only trials with correct responses were included in the RT analysis, so an additional 16.1% of the data were removed. In the neutral condition, monolingual children responded with 94% accuracy, and bilinguals with 93% accuracy. In the incongruent condition, monolinguals gave 94.8% correct responses, and bilinguals 91.6%. In the incongruent condition monolinguals answered correctly 53.3% of the time, and bilingual children 58.2% of the time. One monolingual child was excluded from the analysis since he gave 100% of incorrect answers in the incongruent condition. Mean and median RT in ms for each condition with SD and range (i.e., Min and Max) values for each group are presented in Table 42 below.

Table 42

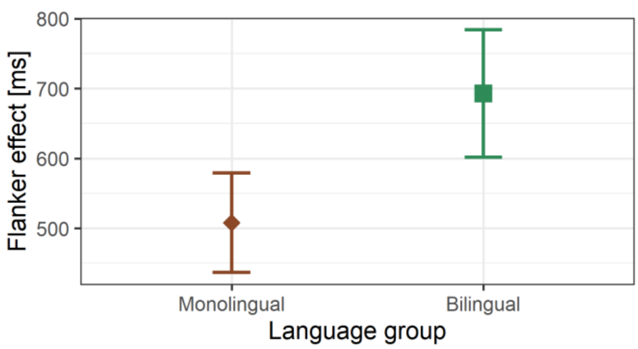
Descriptive values of 4-year-olds in the Flanker tasks

		Neutral	Congruent	Incongruent
Monolinguals				
	Mean	1540	1571	2079
	Median	1505	1566	2040
	SD	273	296	464
	Min	1083	1055	1088
	Max	2255	2326	3022
Bilinguals				
	Mean	1732	1732	2425
	Median	1710	1773	2422
	SD	235	248	512
	Min	1331	1232	1522
	Max	2228	2297	3584

Figure 32 shows the Flanker effect (RT of incongruent minus congruent trials) as a measure of inhibition. A smaller Flanker effect is considered to reflect a better developed inhibition ability.

Figure 32

Mean Flanker effect for monolingual and bilingual 4-year-olds



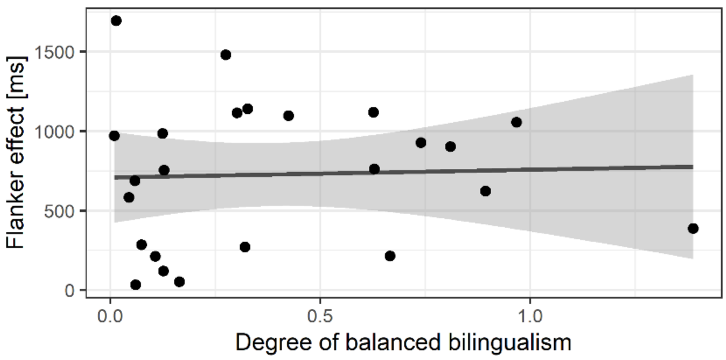
Note. Monolingual children are represented in brown and bilingual children in green. Error bars represents ± 1 SE of between-subject variance.

The Flanker effect was 508 ms for monolinguals and 693 ms for bilinguals. Since the data were normally distributed (W = .976, p = .370), an unpaired t-test was computed. The performance by monolingual and bilingual children did not differ statistically from each other ($t(47)$ = -1.597, p = .117).

In order to investigate the relationship between the Flanker effect and the degree of balanced bilingualism in children, the children's ratio scores of TROG were calculated. As in Study 1, bilingual children with a ratio score of 1 ± 0.15 were considered as balanced. Ten bilingual children were included in the balanced bilingual subgroup, 14 children were included in the unbalanced subgroup and two other children were excluded from this analysis due to missing data (one child did not perform the TROG in Italian, and a second child neither in German nor in Italian, see Appendix E.26 for all individual data). The ratio scores were then converted[12] (analogously to Study 1) on a scale of 0 to 1, in which 0 represents a balanced degree of bilingualism between German and Italian and 1 a language dominance in German or Italian (the closer the values are to 0 the more well-balanced the two languages are). Figure 33, below, represents the correlation graph between the Flanker effect and the degree of balanced bilingualism in children.

Figure 33

Correlation plot Flanker effect and degree of balanced bilingualism in 4-year-olds



Note. Each dot represents one child. The shaded area indicates ±1 SE of between-subject variance.

The Spearman non-parametric correlation test revealed no significant correlation between the Flanker effect and the degree of balanced bilingualism (rho = .109, p = .612)[13].

[12] The value of 1 was subtracted from all ratio scores and the results were then converted into absolute values, as their distance from 0.

[13] A group comparison between balanced and unbalanced bilinguals in the Flanker task was also calculated. The mean RT of the Flanker effect was 631 ms for balanced bilinguals and 795 ms for unbalanced bilinguals. A statistical comparison between the two groups displayed no differenced (unpaired t-test $t(17)$ = 0.834, p = .415).

Concluding, both language groups performed similarly during the Flanker task and a more even balance had no effect on the performance by bilingual children.

Task-switching paradigm

In the task-switching paradigm, practice items were also excluded from the analysis. None of the 4-year-olds attained a response time of below 200 ms. For the RT analysis, only the trials with correct responses were included. Thus, 18.8% of the data had to be excluded. Accuracy data for monolingual children in the non-switch blocks reached 89.2% and for bilinguals 92.5%. Accuracy for non-switch trials of the switch blocks reached 79.2% in the monolingual group and 81.4% in the bilingual group. For switch trials of the switch blocks, monolinguals attained 70.8% accuracy and bilinguals 77.7%. Mean and median RT in ms for each condition with SD and range values for each group are reported in the Table 43 below.

Table 43

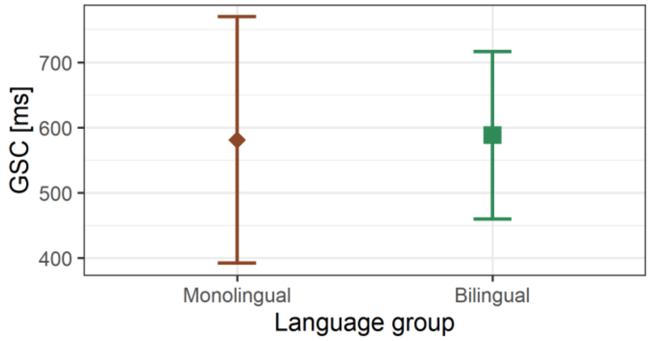
Descriptive values of 4-year-olds in the task-switching paradigm

	Non-switch blocks	Switch blocks non-switch trials	Switch blocks switch trials
Monolinguals			
Mean	1673	2254	2245
Median	1556	2090	2046
SD	564	876	911
Min	914	1496	1444
Max	3653	5478	5593
Bilinguals			
Mean	1636	2224	2291
Median	1554	2076	2078
SD	405	671	754
Min	998	1181	1147
Max	2599	4119	4009

Figure 34 displays the mean GSC (RT of non-switch trials in switch blocks minus the RT of non-switch trials in the non-switch blocks), the dependent variable used as measurement for switching attention in children.

Figure 34

Mean GSC for monolingual and bilingual 4-year-olds



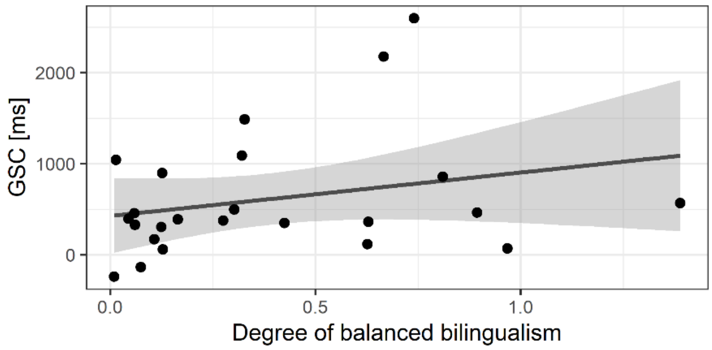
Note. Error bars represents ± 1 SE of between-subject variance.

The monolinguals' GSC was 581 ms and the bilinguals' GSC was 588 ms. Since data were not normally distributed (W = .848, p < .001), a Mann-Whitney-U-Test was employed and revealed no significant differences between the two groups (W = 323, p = .78).

A correlation analysis was also applied between the GSC and the degree of balanced bilingualism in children. This is illustrated in the next Figure.

Figure 35

Correlation plot GSC and degree of balanced bilingualism in 4-year-olds



Note. Each dot represents one child. The shaded area indicates ±1 SE of between-subject variance.

The Spearman non-parametric correlation test showed no significant correlation between the GSC and the degree of balanced bilingualism (*rho* = .319, p = .129)[14].

[14] A group comparison between balanced and unbalanced bilinguals in the task-switching paradigm was also calculated. The mean RT of the Flanker effect was 326 ms for balanced bilinguals and 813 ms for unbalanced bilinguals. Despite this large numerical difference, the Mann-Whitney-U-Test (W = 38, p = .061) only revealed a trend for a larger effect in unbalanced compared to balanced bilinguals.

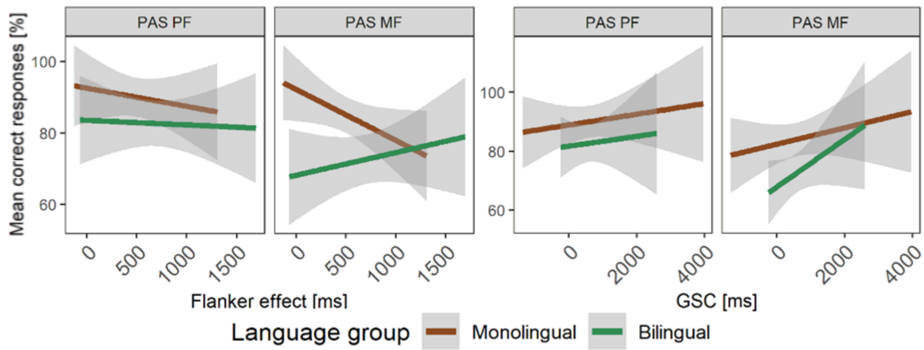
To summarize, these results showed no differences between monolingual and bilingual children in their performance during the Flanker task or the task-switching paradigm. In addition, children's language balance, based on the TROG ratio score performed in German and Italian, did not affect these results.

6.2.5 Passive sentence comprehension and executive functions

Similarly to Study 1, further analyses were run to investigate whether the Flanker effect or the GSC affected the comprehension of passive sentences. Figure 36, below, presents how the mean proportion of correct responses in the sentence comprehension task differs as a results of variation in EF skills.

Figure 36

Linear regression lines reflecting the relation between correct responses for passive sentences and the Flanker effect and the GSC in 4-year-olds separated by condition



Note. The x-axes report the time in ms for the Flanker effect and the GSC. Shaded areas indicate ±1 SE of between-subject variances.

Repeated contrast coding was used for condition: PAS PF was subtracted from PAS MF, i.e., a negative estimate indicates a higher number of correct responses for PAS PF, and a positive estimate a higher number of correct responses for PAS MF. First monolinguals were set to be in the intercept, and next bilinguals were. Three random components were used: participants and items as random intercepts and the individual adjustment of condition for each participant as random slope. The Flanker effect and GSC were added as continuous predictors. The model's parameters are reported in the Table 44 (see Appendix E.27).

Table 44

Results of the covariates analysis for EFs and the 4-year-old monolinguals in the intercept

Fixed effect	**Estimate**	**SE**	***z*-value**	***p*-value**
Intercept (Mono, grand mean)	2.067	0.200	10.312	<.001***
Bili (grand mean)	-0.690	0.249	-2.774	=.006**
PAS MF-PAS PF (Mono)	-0.708	0.354	-2.000	=.046*
Flanker effect (Mono)	-0.939	0.538	-1.745	=.081±
GSC (Mono)	0.265	0.229	1.156	=.248
Bili x PAS MF-PAS PF	-0.044	0.448	-0.099	=.921
Bili x Flanker effect	1.048	0.630	1.663	=.096±
Bili x GSC	0.064	0.348	0.182	=.855
PAS MF-PAS PF x Flanker effect	-0.402	0.981	-0.410	=.682
PAS MF-PAS PF x GSC	-0.029	0.428	-0.068	=.946
Bili x PAS MF-PAS PF x Flanker effect	0.905	1.137	0.796	=.426
Bili x PAS MF-PAS PF x GSC	0.359	0.633	0.568	=.570

‡ *p* <.1, * *p* <.05, ** *p* <.01, *** *p* < .001

Fixed effects of conditions and groups are not discussed here again. Neither the Flanker effect nor the GSC affected the comprehension of passive sentences in the monolingual group (all *p* > .081).

The second model included bilinguals in the intercept. The other parameters were coded as in the previous model (see Appendix E.28 for model output). Parameters of this analysis are displayed in the Table 45, below.

Table 45

Results of the covariates analysis for executive functions and the 4-year-old bilinguals in the intercept

Fixed effect	**Estimate**	**SE**	***z*-value**	***p*-value**
Intercept (Bili, grand mean)	1.378	0.165	8.371	<.001***
Mono (grand mean)	0.690	0.249	2.774	=.006**
PAS MF-PAS PF (Bili)	-0.752	0.275	-2.736	=-006**

Fixed effect	**Estimate**	**SE**	***z*-value**	***p*-value**
Flanker effect (Bili)	0.110	0.328	0.335	=.738
GSC (Bili)	0.329	0.262	1.255	=.209
Mono x PAS MF-PAS PF	0.044	0.448	0.098	=.922
Mono x Flanker effect	-1.049	0.630	-1.664	=.096‡
Mono x GSC	-0.063	0.348	-0.179	=.858
PAS MF-PAS PF x Flanker effect	0.502	0.579	0.866	=.386
PAS MF-PAS PF x GSC	0.329	0.467	0.703	=.482
Mono x PAS MF-PAS PF x Flanker effect	-0.905	1.137	-0.796	=.426
Mono x PAS MF-PAS PF x GSC	-0.358	0.633	-0.566	=.571

‡ p <.1, ** p <.01, *** p < .001

Hence, neither the Flanker effect nor GSC are predictors for the understanding of passive sentences in bilingual children (all p > .209).

To summarize, the EF abilities measured cannot be not associated with the accuracy in the comprehension of passives in these two groups.

6.2.6 Linguistic, cognitive and social predictors

Two different covariate analyses are reported in this section. The first analysis investigated which predictors retrieved from the bilingual language background questionnaire (i.e., cumulative input and output, parents' education and the number of older school-aged siblings, see also Section 6.2.1) and from the monolingual language background questionnaire (i.e., parents' education and the number of older school-aged siblings) might be associated with the children's performance on passive sentence comprehension. The second analysis concerned the monolinguals' and bilinguals' data, which had been extracted from the standardized tests (language proficiency, memory capacity and reasoning ability). The aim was to investigate whether these variables influence the comprehension of passives in monolingual and bilingual children.

Active sentence conditions were not considered in these analyses, since the focus of the study was on the performance of children in passive sentences. Before proceeding with the first covariate analysis for the bilingual group, correlations between the variables were calculated using Spearman's rho and reported in Table 46, below.

Table 46

Correlation coefficients between variables from the language background questionnaire in 4-year-old bilinguals

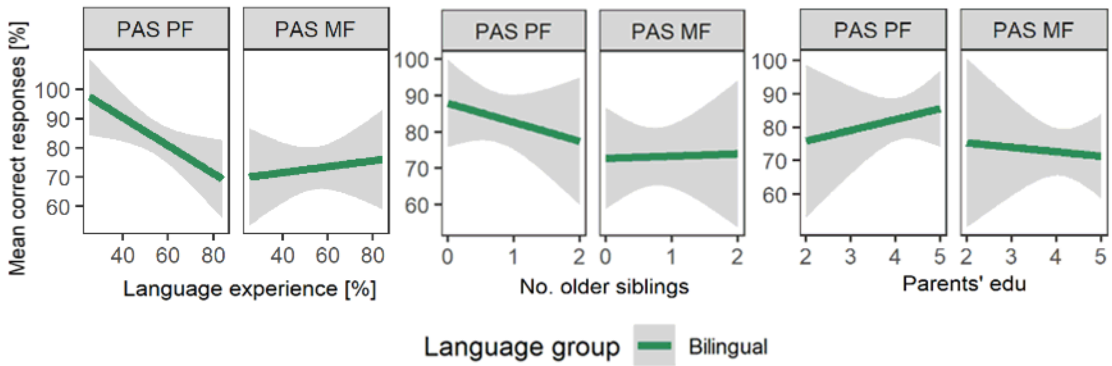
	Input	**Output**	**Older siblings**	**Parents' edu**
Input	1			
Output	0.689***	1		
Older siblings	0.209	0.358	1	
Parents' edu	-0.294	-0.436*	-0.084	1

* $p < .01$, *** $p < .001$

Due to high correlation between German language input and output, the new composite variable language experience was created. As in Study 1, the language experience variable is the mean between the variables of language input and language output (calculated as described in Section 5.2.1).

Figure 37

Linear regression lines reflecting the relation between correct responses for passives and the effect of language experience, number of older siblings and parents' education in bilingual 4-year-olds



Note. The number of older siblings is indicated on the x-axis: the zero represents no older siblings, i.e., only children or children with younger siblings only. The x-axis of the parents' education graph shows a mean between the educational level of both parents on a scale in which 5 is the highest degree of education. Shaded areas indicate ±1 SE of between-subject variance.

A LMM was run, in which a repeated contrast for the passive condition was applied: the intercept reflects the average performance in both passive conditions. All continuous predictors were centered to the group's average to ensure interpretation of the intercept. Language experience, the number of older siblings (no. older siblings) and parents' education (parents' edu) were added into the model as numeric predictors. Three random components were included in the models: two random intercepts for participants and items and a random slope, the effect of condition for each participant (see Appendix E.29 for the complete model output). Table 47 reports the parameters of this model.

Table 47

Results of the covariates analysis retrieved from the language background questionnaire for 4-year-old bilingual children

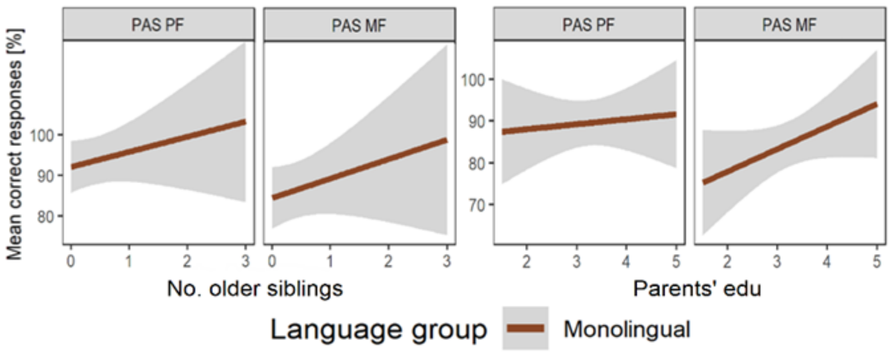
Fixed effect	**Estimate**	**SE**	***z*-value**	***p*-value**
Intercept (grand mean)	1.535	0.187	8.195	<.001***
PAS MF-PAS PF	-0.873	0.319	-2.740	=.006**
Lang. experience	-0.021	0.013	-1.637	=.102
No. older siblings	-0.050	0.344	-0.146	=.884
Parents' edu	-0.120	0.285	-0.420	=.674
PAS MF-PAS PF x lang. experience	0.033	0.021	1.591	=.112
PAS MF-PAS PF x no. older siblings	0.241	0.583	0.414	=.679
PAS MF-PAS PF x parents' edu	0.095	0.483	0.196	=.844

* p <.05, *** p < .001

Accuracy data are not discussed again. Regarding the predictors, none affected bilingual children's performance on passive sentences across conditions (all p > .102). Regarding the monolingual group, a similar analysis was performed, in order to investigate whether there was an influence of the number of older siblings or of the parents' education on their understanding of passive sentences (see Figure 38).

Figure 38

Linear regression lines reflecting the relation between correct responses for passive sentences and the effect of number of older siblings and parents' education in monolingual 4-year-olds



Note. The number of older siblings is indicated on the x-axis: the zero represents no older siblings, i.e., only children or children with younger siblings only. The x-axis of the parents' education shows a mean between the educational level of both parents on a scale in which 5 is the highest degree of education. Shaded areas indicate ±1 SE of between-subject variances.

The model specification was identical to the previous model, with the exception that the predictor of language experience is not available in this case (see Appendix E.30 for model output). Table 48 reports these parameters.

Table 48

Results of the covariates analysis retrieved from the language background questionnaire for 4-year-old monolingual children

Fixed effect	Estimate	SE	*z*-value	*p*-value
Intercept (grand mean)	6.387	5207	0.001	=.999
PAS MF-PAS PF	-7.823	10414	-0.001	=.999
No older sibling	10.01	14134	0.001	=.999
Parents' edu	0.313	0.368	0.850	=.395
PAS MF-PAS PF x no. older siblings	-18.17	28267	-0.001	=.999
PAS MF-PAS PF x parents' edu	0.214	0.836	0.256	=.798

As shown above, neither the number of older siblings nor the parents' education modified the accuracy in comprehension of passive sentences in the monolingual group (all $p > .395$).

As a second covariate analysis, the influence of a general language knowledge and cognitive measures was analyzed to establish the performance of monolingual and bilingual children in their comprehension of passive sentences. Spearman correlation coefficients were calculated between the language and cognitive raw scores, which the children attained in the various standardized tests.

Table 49

Correlation coefficients between raw scores from the standardized tests in monolingual and bilingual 4-year-olds

	German PPVT	**German TROG**	**Visual memory**	**Digit span**	**CPM**
German PPVT	1				
German TROG	0.295*	1			
Visual memory	0.209	0.373**	1		
Digit span	0.077	0.200	0.453***	1	
CPM	-0.014	0.360**	0.249	0.406**	1

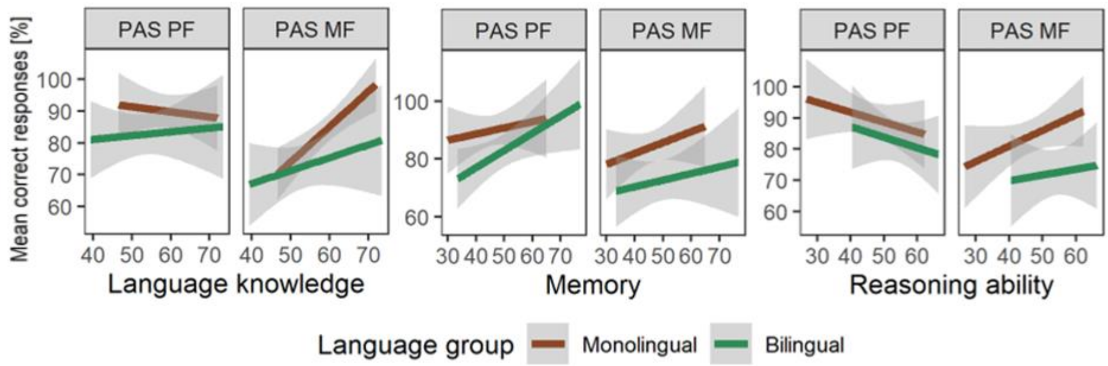
$^{*}p$ <.05, ** p <.01, *** p < .001

As in Study 1, a composite variable for language knowledge (mean T-scores of the German PPVT and the German TROG) and a variable for memory (mean T-scores of the visual memory and digit span tests) were created. T-scores of the CPM were used as measure of reasoning ability and additionally included in the model.

Thus, for the second covariates analysis, the two levels of the factor condition (PAS MF-PASPF) were compared with repeated contrast (as before). Language group was coded as treatment contrast in bilingual children as a reference group in the intercept. The three random components reflected the previous analysis. One bilingual child was excluded from the model on account of missing value in the CPM test. Figure 39, below, shows the linear regression lines for monolinguals and bilinguals between mean accuracy (y-axes) in passive sentences and the three potential predictors: language knowledge, memory and reasoning ability (x-axes).

Figure 39

Linear regression lines reflecting the relation between correct responses for passive sentences and the effect of language knowledge, memory and reasoning ability in 4-year-olds



Note. Brown lines represent monolinguals and green lines bilinguals. Units of measurement on the x-axes are T-scores. Shaded areas indicate ±1 SE of between-subject variance.

Table 50 below presents the parameters of this covariate analysis (see Appendix E.31 for complete model output).

Table 50

Results of the covariates analysis with the standardized tests and 4-year-old bilinguals in the intercept

Fixed effect	**Estimate**	**SE**	***z*-value**	***p*-value**
Intercept (Bili, grand mean)	1.536	0.196	7.829	<.001***
Mono (grand mean)	0.491	0.285	1.724	=.085‡
PAS MF-PAS PF (Bili)	-0.911	0.347	-2.628	=.009**
Lang knowledge (Bili)	0.026	0.017	1.498	=.134
Memory (Bili)	0.047	0.017	2.822	=.005**
Reas ability (Bili)	-0.039	0.022	-1.797	=.072‡
Mono x PAS MF-PAS PF	0.242	0.533	0.454	=.650
Mono x lang knowledge	0.007	0.030	0.230	=.818
Mono x memory	-0.023	0.029	-0.775	=.438
Mono x reas ability	0.023	0.031	0.756	=.450
PAS MF-PAS PF x lang knowledge (Bili)	-0.004	0.031	-0.144	=.886
PAS MF-PAS PF x memory (Bili)	-0.069	0.031	-2.193	=.028*
PAS MF-PAS PF x reas ability (Bili)	0.065	0.039	1.659	=.097‡

Mono x PAS MF-PAS PF x lang knowledge	0.113	0.057	1.995	=.046*
Mono x PAS MF-PAS PF x memory	0.019	0.055	0.334	=.731
Mono x PAS MF-PAS PF x reas ability	0.002	0.058	0.035	=.972

‡ *p* <.1, * *p* <.05, ** *p* <.01, *** *p* < .001

Across conditions, memory had a positive effect on the bilinguals' comprehension of passive sentences (*p* =.005). A significant interaction of PAS MF-PAS PF x memory with a negative estimate indicated a stronger effect in PAS PF sentences (*p* =.028). The significant three-way interaction Mono x PAS MF-PAS PF x lang knowledge indicated a different effect of language knowledge on the monolingual group between the passive conditions, in comparison to the bilingual group (*p* =.046).

Table 51 shows the parameters of the model with the monolingual in the intercept (see Appendix E.32 for the complete model outputs).

Table 51

Results of the covariates analysis with the standardized tests and 4-year-old monolinguals in the intercept

Fixed effect	**Estimate**	**SE**	***z*-value**	***p*-value**
Intercept (Mono, grand mean)	2.027	0.222	9.124	<.001***
Bili (grand mean)	-0.491	0.285	-1.723	=.085‡
PAS MF-PAS PF (Mono)	-0.666	0.403	-1.653	=.098‡
Lang knowledge (Mono)	0.033	0.025	1.308	=.191
Memory (Mono)	0.025	0.024	1.016	=.310
Reas ability (Mono)	-0.015	0.022	-0.678	=.498
Bili x PAS MF-PAS PF	-0.246	0.533	-0.462	=.644
Bili x lang knowledge	-0.007	0.030	-0.231	=.818
Bili x memory	0.023	0.029	0.772	=.440
Bili x reas ability	-0.023	0.031	-0.760	=.448
PAS MF-PAS PF x lang knowledge (Mono)	0.108	0.047	2.288	=.022*
PAS MF-PAS PF x memory (Mono)	-0.050	0.045	-1.100	=.272

PAS MF-PAS PF x reas ability (Mono)	0.067	0.043	1.570	=.116
Bili x PAS MF-PAS PF x lang knowledge	-0.113	0.057	-1.992	=.046*
Bili PAS MF-PAS PF x memory	-0.019	0.055	-0.343	=.731
Bili x PAS MF-PAS PF x reas ability	-0.002	0.058	-0.040	=.967

‡ p <.1, * p <.05, ** p <.01, *** p < .001

Across conditions, neither language knowledge, memory nor reasoning ability influenced monolinguals' understanding of passive sentences (all p > .191). However, a significant interaction of PAS MF-PAS PF x lang knowledge with a positive estimate indicated a stronger effect in PAS MF than PAS PF sentences (p =.022). The three-way interaction Bili x PAS MF-PAS PF x lang knowledge reflected the result in the previous model.

In summary, the only variable to affect the comprehension of passive sentences in the bilingual group is memory: the higher the memory score, the higher the children's performance in the sentence comprehension task.

6.3 Discussion

The aim of Study 2 was to investigate how German monolingual and German-Italian bilingual children perform in the comprehension and processing of passive sentences, in which the subject was placed in the prefield or in the middlefield. In addition to response accuracy, eye-gaze and pupil data were also analyzed, as measures of sentence processing. A further central investigation of this study was to analyze whether EF abilities, as well as language-external variables and scores of receptive vocabulary and grammar, visual and short-term memory and reasoning ability influenced children's comprehension of passive sentences.

6.3.1 Sentence comprehension and processing

The RQ1 asked whether monolingual and bilingual children performed differently in their comprehension and processing of passives, depending on the subject position in the prefield or middlefield. As a result, the analysis of the accuracy data demonstrated that all the children performed above chance in all sentence conditions. The accuracy data of the monolingual group showed no significant differences in the comprehension of the two passive sentence conditions, whereas bilingual children performed better in the passive sentences in which the subject was in the prefield rather than in the middlefield. Huang et al. (2013) first hypothesized that when the NP1 in a sentence does not occur before the voice cue, children's thematic roles assignment is more prone to wait for the disambiguating cue, which weakens their subject-/agent-first bias. Their findings with Mandarin-Chinese 5-year-olds proved that children performed better when the passive cue appeared before the subject. In contrast, Ehrenhofer's (2018) found that German 5-year-olds performed well in passive comprehension, regardless of whether the passive voice cue preceded or followed the subject. The accuracy data of German 4-year-olds in the current study replicated Ehrenhofer's findings (2018), as no difference in performance was established between the two passive sentence conditions. Following Aschermann et al. (2004) and Armon-Lotem et al. (2016), it is suggested that the V2 properties in German mean that children do not rely consistently on word order to identify the subject of the sentence, contrary to what happens in other languages such as English.

During the processing of passive sentences, it has been found that monolingual children looked more to the target during active sentences with the subject in the prefield than passive sentences with the subject in the prefield, after having heard the voice cue *wurde*. However, this difference was only present in one time window. This may show that monolingual children process active and passive sentences similarly, after a first preference for active sentences. Within the bilingual group, differences between sentence conditions were realized after sentence offset. During the first time window of silence, bilinguals looked more to the target during active sentences with the subject in the middlefield than during passive sentences with the subject in the middlefield. During the second time window of silence, namely 750 ms after sentence offset, the bilinguals' proportion of looks to the target was smaller during passives with the subject in the middlefield than during passives with the subject in the prefield, as well as during the two active sentence conditions. During the third time window of silence after sentence offset, namely 1500 ms after sentence offset, bilinguals looked more to the target during active sentences with the subject in the prefield, than during passive sentences with the subject in the middlefield. The higher proportion of target looks by bilingual children during the processing of active than passive sentences after sentence offset might indicate a better recognition of the auxiliary verb for active sentences than in passive sentences, independently of the subject position. The similar performance by monolingual and bilingual children during passive sentence processing contradicts the results of Cristante (2016), whose study found that eL2 Turkish-German 7-year-olds showed a different looking behavior than their monolingual peers: the target looking of eL2 children was only above chance after sentence offset, whereas monolingual children were above chance immediately after the voice cue *wurde*. She suggested that their length of exposure to German might have played a role. Instead, in the present study, the length of exposure to German of the bilingual children corresponds to that of the monolingual children, which might primarily explain the diverging results.

Turning to pupillometry data, considerable differences were found within the two groups. While monolinguals had larger and more pronounced fluctuations in the different conditions, the pattern in bilinguals was quite uniform. Monolingual children showed larger pupil dilation during passives with the subject in the prefield rather than in the middlefield during the first time window (from 800 ms to 1800 ms, i.e., after the onset of the voice cue *wurde*), and larger pupil dilation during passives with the subject in the middlefield rather than in the prefield (from 1800 ms to 2800 ms). Hence, pupillometry data revealed patterns in the monolingual group that were not detected by eye-gaze data. It could be assumed that the voice cue *wurde* before the subject helped children to recognize passive sentences during the first time window, and that this preference was not detectable during the second time window, due to frequency factors, and the preference deriving from that, for the subject to be in the prefield. The bilingual group showed larger pupil dilation during passives with the subject in the prefield than in the middlefield after having heard the voice cue *wurde*, a pattern which was displayed during both time windows, from 800 ms to 2800 ms after sentence onset. This finding is not in line with accuracy and eye-gaze data that showed a preference for passives with the subject in the prefield rather than in the middlefield (even though this preference was only manifested during the second time window of silence after sentence offset in the eye-gaze data). The pupillometry data from the bilingual group correspond to the proposal of Huang et al. (2013): when the sentence subject occurs after the disambiguating cue (in this case the passive voice cue), the children's subject-/agent-first bias is weakened. However, it remains unclear as to why eye-gaze and pupil data lead to contradicting results, and therefore there is a need for further investigation.

With regard to RQ2: the question was posed as to whether bilingual children manifested any differences in their comprehension and processing of passive structures, in comparison to the monolingual group. The accuracy data showed that the performance of monolingual and bilingual children in their comprehension of passives with the subject in the prefield was similar; however, in the comprehension of passives with the subject in the middlefield, monolinguals outperformed bilinguals. There are two possible reasons why bilingual children perform more accurately in passive sentences in which the subject is in the prefield rather than the middlefield: first, frequency factors and second, cross-linguistic influences. Engel (1974) found 51% subject-initial and 35% adverbial-initial main clauses in a corpus of colloquial spoken German, as well as in written German. Bohnacker (2007) reported similar results, based on a colloquial German corpus recorded in the late 1990s: 52% subject-initial and 37% adverbial-initial main clauses. German maintains the tendency to initiate declarative sentences with a subject (Bohnacker & Rosén, 2008) even though the production rate of adverbial-initial German main clauses is high. On the other hand, cross-linguistic influences might also have played a role in bilinguals' performance with passive sentences when the subject is in the middlefield.

Word order in which the subject is postverbal (in a broad focus context) is not possible in Italian (see also Section 6). Fronting an adjunct requires the subject to be in the preverbal position, thus maintaining the Italian SVO word order. This might have led children to achieve better results in the passive sentence condition that overlaps between the two languages (i.e., passive sentences with the subject in the prefield). However, eye-gaze data showed that the proportion of looks to the target by monolingual and bilingual children did not differ during the processing of the two passive conditions.

The RQ3 was designed to explore whether there was a difference in the performance of children between online and offline tasks. A further analysis of eye-gaze data was run, with the factor accuracy added into the model. The findings showed that both bilingual and monolingual children looked more to the target when giving correct responses rather than incorrect responses across sentence conditions. This signifies that eye-gaze data predicted children's responses. In fact, monolingual children did not show any differences between sentence conditions in the offline task, and their looking behavior reflected the same pattern. Bilingual's children eye-gaze data demonstrated that it was only in one time window of silence after sentence offset that the proportion of looks to the target was smaller during passive sentences with the subject in the middlefield than during passives with the subject in the prefield. This might reflect the moment in which bilingual children gave an answer in the sentence comprehension task, when the effort required to perform in the offline task turned out to be visible in the less accurate sentence condition. However, this suggestion needs to be supported by further data analyses. After the addition of the factor accuracy into the pupillometry data model, the bilingual group demonstrated larger pupil dilation when giving incorrect responses than in correct responses across conditions, whereas in monolingual children, it was during correct responses that pupil dilation was larger across conditions. The larger pupil dilation of bilingual children in incorrect responses may signal that these children have detected the passive cues, even though responses in the offline task were incorrect. The larger pupil dilation in monolingual children when giving correct responses predicted children's responses.

6.3.2 Sentence comprehension and executive functions

As in Study 1, the RQ4 aimed to explore whether EF abilities have an effect on children's performance in the sentence comprehension tasks. First, the performance of children in the EF tasks was analyzed. The findings provided evidence that the performance of bilingual and monolingual children was comparable. To investigate whether the degree of balanced bilingualism played a role on bilinguals' performance, children were divided into a balanced and an unbalanced subgroup on the basis of their ratio of German and Italian TROG score (Bishop, 2003). It was found that the more balanced bilingual children did not display any differences in their performance of EF tasks compared to the unbalanced bilingual children. These findings contradict the hypothesis of the existence of a bilingual cognitive advantage. It is suggested that variables affecting the bilingual language experience should be systematically recorded; furthermore, it is also important to establish criteria for defining language balance, as well as the description of experimental tasks employed to measure EFs in children, in order to be able to replicate previous results (see also Section 5.3.2 for further discussion).

Furthermore, no correlation was found between the mean percentage of correct responses in the sentence comprehension task and the EF tasks. The reason why EFs do not influence passive comprehension might be that children already feature the necessary linguistic competences to process this syntactic structure. Therefore, they may not need to be supported by EF abilities during sentence comprehension.

6.3.3 Sentence comprehension and linguistic, cognitive and social predictors

The role that language-external factors may have on sentence comprehension was also taken into consideration. The RQ5 asked whether the performance of monolingual and bilingual children was affected by their parents' education and by the number of school-aged older siblings. The language experience in the bilingual group was also calculated and considered as a potential predictor for bilinguals' passive sentence comprehension. Initially, a comparison between the two groups showed that the bilingual group included higher parents' education, as well as a larger number of older school-aged siblings, than the monolingual group did. However, these differences did not influence children's comprehension of passive sentences in any way: no correlation of these variables with the accuracy score was found in either group. Nor was the language experience variable a significant predictor for bilinguals' comprehension of passive sentences.

Finally, the language knowledge, memory capacity and reasoning abilities of all the children were measured. The RQ6 asked whether these factors could be predictors for the comprehension of passive sentences. Regarding the standardized tests, monolingual children outperformed bilingual children in the German PPVT receptive vocabulary test, while bilinguals outperformed monolinguals in the CPM test, which measures reasoning ability. Previous studies had also reported a higher performance by monolingual children in receptive vocabulary tests over bilingual children (e.g., Calvo & Bialystok, 2014; Chiat & Polišenská, 2016). The higher performance by bilinguals in the CPM task over monolinguals had equally been reported in previous studies (e.g., Tsimpli et al., 2020). The variables of language knowledge (measured by means of a receptive vocabulary and grammar tasks), memory capacity (measured with a visual and verbal short-term memory), and reasoning ability (measured by the CPM task) were added into a model as predictors for children's understanding of passive sentences.

Memory capacity was the only predictor to influence the bilingual children's performance in passive sentences (and not that of the monolingual children): the higher the memory capacity appeared to be, the better the bilingual children understood passive sentences. This result is in line with the correlation found in Study 1 between the accuracy in OVS sentences and memory capacity in the bilingual group. As proposed for Study 1, a greater demand for working memory might be caused by the storage of information during complex sentence processing.

6.3.4 Summary

In conclusion, German monolingual and German-Italian simultaneous bilingual 4;6-year-olds did not display any difficulties in their comprehension of German passive sentences. Two types of passives were used for this study: passive sentences with the subject in the prefield, and passive sentences with the subject in the middlefield occurring after the auxiliary verb. The monolingual children performed similarly during their comprehension and processing of these two sentence conditions, whereas the bilingual children performed more accurately in their comprehension of passives with the subject in the prefield, rather than in the middlefield. The bilinguals' eye-gaze data reflected this finding as early as one time window of silence after sentence offset. When children were prompted to give answers in the offline task, more looks to the target were detected during passives with the subject in the prefield than with those in the middlefield. It was proposed that the difficulties that arose in the offline task were more visible in passives with the subject after the auxiliary verb, because this structure does not overlap between the bilinguals' two languages and because of frequency factors. Inhibition and switching abilities did not play a role in the children's comprehension of passives. The bilingual children's comprehension of German passive sentences was affected by their memory capacity, leading bilingual children to a higher accuracy score in the sentence comprehension task.

Part III Conclusions

7 Conclusions

The main goal of this dissertation was to understand how monolingual and bilingual children use morpho-syntactic cues to assign thematic roles in non-canonical sentences, as well as to investigate whether EFs (inhibition and switching abilities) influence children's language comprehension. The comprehension and processing of German OVS and passives was tested among German monolingual and German-Italian simultaneous bilingual children by implementing offline and online methods. These two non-canonical structures are known challenges in sentence comprehension: driven by subject/agent-first-bias, listeners may assume the object/patient in OVS sentences to be the subject/agent, and the subject in passive sentences to be the agent rather than the patient. Based on the assumption of incremental processing, these biases would lead to an initial misinterpretation of the sentence, and a correct interpretation would require syntactic and thematic revision.

It was found that German monolingual 5;8-year-old children made use of the second disambiguating nominative cue on the NP2 to correctly assign thematic roles in OVS sentences in which the first disambiguating accusative cue was on the NP1, while German-Italian simultaneous 5;8-year-old bilingual children did not do so. The chance level performance of bilingual children in the same sentence condition could be interpreted as an indicator that some children of that age have already acquired the disambiguating case-marking cue on the NP1 while others have not. However, it could also be presumed that some children showed chance level performance at an individual level. There needs to be further analyses for more conclusive results. However, when OVS sentences contained a disambiguating nominative cue on the NP2, providing an ambiguous accusative case-marking cue on the NP1, both language groups initially assigned the role of agent to the NP1. It was only after having heard the unambiguous nominative case-marking cue on the NP2 that both groups started a revision analysis, a finding which was supported by the eye-gaze data. The length of the ambiguous region proved to have an effect on the comprehension and processing in both groups: the longer the ambiguous region lasted, the greater the difficulty the children had in the assignment of thematic roles. Between groups, the monolingual children outperformed the bilingual children during the comprehension of both types of OVS sentences.

Pupil data in the monolingual group revealed no pupil size changes between conditions, while the bilingual group showed a larger pupil dilation during the processing of OVS sentences with the disambiguating cue on NP2, than when it was on NP1. It was suggested that pupil size changes were not detected in the monolingual group because they are at a different stage of case-marking acquisition (as also supported by accuracy and eye-gaze data) compared to the bilingual group's stage. However, pupillometry data also revealed that bilingual children had a larger pupil dilation when giving incorrect responses than in their correct responses across condition (this result was not reflected in the eye-gaze data). This finding may suggest that children have detected the disambiguating cue, despite giving incorrect responses in the sentence comprehension task. If this is the case, pupillometry data might be a more finely grained measuring method than eye-gaze data, since it plots a children's sensitivity for the disambiguating case-marking cues, even in incorrect responses.

Study 2 demonstrated that passive sentences were well comprehended by all 4;6-year-old children. Monolingual children did not show any differences in their evaluation of whether the subject was in the prefield or in the middlefield, after the voice cue (auxiliary *wurde*). By contrast, bilingual children showed more accurate performance in passives in which the subject was in the prefield than subjects in the middlefield. A difference in looking behavior between the two groups was not observed. The monolinguals' looking behavior was comparable between the two passive conditions. Before sentence offset the looking behavior of bilingual children did not differ between the two passive conditions, after sentence offset they looked more to the target during passives with the subject in the prefield than during those in the middlefield. This can be explained in terms of major effort required by the offline task, more visible in the non-overlapping sentence structure in the two languages of bilinguals. Therefore, it is argued that cross-linguistic influences played a role in bilingual performance. By contrast, a frequency effect may also have had an impact on the bilinguals' performance, since sentences with the subject as NP1 are the most frequent structure in German main clauses. Furthermore, eye-gaze data predicted children's responses.

The bilinguals' pupil data revealed that the processing of passives with the subject in the prefield caused larger dilation than the processing of passives with the subject in the middlefiled, whilst in the monolingual group this effect was only found during the first time window. During the second time window of analysis (i.e., during the *by*-phrase after sentence offset), they displayed larger pupil dilation during the processing of passives with the subject in the middlefield than they did in the condition with the subject in the prefield. It is proposed that the voice cue *wurde* before the subject might help children in the assignment of thematic roles preventing the agent-first bias, and that this preference was abandoned in favor of a subject perceived in the prefield during the second time window. While monolingual children showed larger pupil dilation when giving correct responses, bilingual children displayed the opposite pattern. This may signal that bilingual children have detected the passive cues even though responses in the offline task were incorrect.

It is suggested that OVS sentences might be more difficult to process than passives sentences. During the comprehension and processing of passive sentences, monolingual and bilingual children performed above chance, contrary to the comprehension and processing of OVS sentences, in which only the monolingual children showed an above chance performance when the disambiguating cue was on the NP1. Two main reasons were proposed as an interpretation of these findings: OVS sentences need a syntactic and thematic reanalysis (as a result of the object/patient sentence initial), whereas passives only need a thematic reanalysis (as a result of the subject/patient sentence initial). Additionally, OVS sentences disambiguate through one nominal cue dependent on three factors (namely case, number and gender), whereas passives disambiguate through two cues: the verbal voice and the *by*-phrase (for similar reasons see Cristante & Schimke, 2018).

The other main research area of this dissertation was related to the development of EFs and the role of these on sentence comprehension. First, analyses were performed to investigate whether bilingual children showed a cognitive advantage over their monolingual peers. In order to keep the bilingual groups as homogeneous as possible, only simultaneous bilinguals (i.e., AoO from birth and one parent-one language households) were selected for both studies. Furthermore, as also suggested by de Bruin (2019), standardized objective proficiency measurements were adopted: a degree of balanced bilingualism (i.e., balanced versus unbalanced) was calculated by means of a receptive grammar task (TROG; Bishop, 2003) completed by bilingual children in their two languages. The findings of both studies approved that there was no bilingual cognitive advantage over their monolingual peers, and also no benefits to be derived from more balanced bilingualism. As also reported in the review by Giovannoli et al. (2020), the inconsistent results of studies on the relationship between bilingualism and EFs propel the need for further studies to be undertaken. The current findings should not be taken as evidence against the beneficial effects of the bilingual experience on executive functioning, but rather as a need to describe in more detail the variables of the bilingual experience, such as language use and language switching in different contexts. V. C. M. Gathercole et al. (2014) suggested, for example, that a bilingual cognitive advantage might be more present in L2 learners than in simultaneous bilinguals, since their language knowledge is assumed to be less automatic and thus to require greater control during the process of language switching.

The effect of EFs on sentence comprehension displayed different patterns between groups and linguistic structures. It was hypothesized that advanced EF abilities might help children's sentence processing by overriding the initially preferred sentence interpretation. As suggested in previous studies (e.g., Höhle et al., 2016; Minai et al., 2012; I. Sekerina et al., 2004), children's failure to revise the initial interpretation of garden-path sentences may be related to their yet incompletely developed EF abilities. In Study 1, children's comprehension of OVS sentences was related to their switching ability, but in two different ways. The bilingual group demonstrated the expected negative correlation, namely the higher the accuracy in the sentence comprehension task was, the lower the GSC (again, global switch cost) in the task-switching paradigm, whereas the monolingual group displayed a positive correlation: higher GSC correlates with higher accuracy. The inhibition ability, however, was never related to children's OVS sentence comprehension. It has been proposed that the abilities involved in the task-switching paradigm (i.e., like working memory, attention and goal processing) are more involved in the resolution of the sentence comprehension task than in the Flanker task measurement. A further proposal explaining the differences between the monolingual and bilingual groups was that the EFs are required to support children's sentence comprehension wherever the linguistic structure is difficult to master. This would also explain why in Study 2 no relationship was found between EF tasks and the comprehension of passive sentences. All the children performed well in the comprehension of passives. Memory capacity positively affected bilingual's OVS and passive sentence comprehension. This is in line with previous studies which have suggested that low working memory correlates with low accuracy scores during the comprehension of complex sentences (see e.g., Arosio et al., 2010).

In this dissertation, bilingualism was additionally calculated as a continuous variable in terms of bilingual language experience (i.e., cumulative input and output). This variable was a predictor for bilinguals' performance in OVS sentences: children that received more German language input, and spoke more German, were able to understand OVS sentences more accurately. In other words, more German experience led children to perform better in the German OVS sentence comprehension task. This was not the case for the bilinguals' performance in passive sentences: language experience did not play a role, probably because they had already acquired passive constructions.

In conclusion, this dissertation provides evidence that monolingual children use case-marking cues to assign thematic roles, even though their performance is dependent on the length of the ambiguous region, whereas bilingual children find themselves at an earlier stage of case-marking acquisition. In the comprehension and processing of passives, all children performed well. It is suggested that the better performance of bilinguals in the passive condition in which the subject was placed before the voice cue is due to the overlapping of the syntactic structures in their two languages. Moreover, these studies prove that a bilingual cognitive advantage is not always detectable, and that the relation between EF abilities and sentence comprehension might depend on the tasks applied and on the linguistic competences already mastered.

Notably, these two studies are the first to investigate the processing of German non-canonical sentences in simultaneous bilingual children. This is significant, not only because they contribute to fill the gap in online studies on sentence processing in simultaneous bilingual children, thereby defining some stepstones in bilingual language development. A further reason for their significance is that this research is an attempt to identify different factors that might influence monolingual and bilingual children's sentence comprehension and processing. A simultaneous bilingual group growing up with the same two languages was selected to account for cross-linguistic influence. The criteria used to select the bilingual group and calculate their degree of balanced bilingualism were carefully described to contribute to the bilingual cognitive advantage debate within a specific population. Language-internal and -external factors were considered, in order to investigate which variables influence children's language comprehension. The different methods employed gave a more complete picture of the phenomena. It was proposed that EF abilities support children's non-canonical sentence comprehension when their linguistic structure is not yet fully mastered. Altogether, these findings provide novel evidence in monolingual and bilingual sentence processing, as well as in the role of EFs during non-canonical sentence comprehension.

Appendix

A Language background questionnaires

A.1 Bilingual language background questionnaire

Fragebogen zum sozialen und sprachlichen Hintergrund

Liebe Eltern,

von den Kindern, die an unseren Studien teilnehmen, benötigen wir einige Informationen zu Entwicklung, Familie und Sprache. Alle Angaben werden vertraulich behandelt und nur in anonymisierter Form gespeichert. Bitte nehmen Sie sich ca. 20 Minuten Zeit, um den Fragebogen sorgfältig und aufmerksam auszufüllen. Falls Sie Fragen oder Wünsche haben, schreiben Sie sie bitte auf der letzten Seite (Teil C – Kommentare der Familie) auf.

1. Heutiges Datum: ________________
Tag Monat Jahr

2. Ausgefüllt von: Sorgeberechtigte/r 1 ☐ Beziehung zum Kind bitte nennen: ______________________
Sorgeberechtigte/r 2 ☐ Beziehung zum Kind bitte nennen: ______________________
Andere ☐ (bitte ausführen) ______________________________________

Teil A – Hintergrund

Die folgenden Informationen beziehen sich auf Ihr Kind:

3. Geburtsdatum: ________________
Tag Monat Jahr

4. Erfolgte die Geburt zum errechneten Termin? Ja ☐ Nein ☐
Wie viele Tage früher? ____________
Wie viele Tage später? ____________

5. Geschlecht: weiblich ☐ männlich ☐

6. Kita-/ Schuljahr: ____________________

7. Geburtsland: ____________________

8. Ist das Sehvermögen des Kindes beeinträchtigt?
Nein ☐ Ja ☐ (bitte Auffälligkeit nennen) ______________________________

9. Hatte das Kind Krankheiten, die das Hörvermögen beeinträchtigen? (z.B. Mittelohrentzündungen)
Nein ☐ Ja ☐ (bitte Auffälligkeit nennen) ______________________________
Wann das letzte Mal? Und wie häufig pro Jahr? ______________________

10. Gibt es diagnostizierte chronische Krankheiten?
Nein ☐ Ja ☐ (bitte Auffälligkeit nennen) ______________________________

11. War das Kind jemals in einer sprachtherapeutischen Behandlung?
Nein ☐ Ja ☐ (bitte Auffälligkeit nennen) ______________________________
Von wann bis wann erfolgte die Sprachtherapie? ______________________

Die folgenden Informationen beziehen sich auf die Sorgeberechtigten:

12. Gibt es einen nahen Verwandten des Kindes, der mit dem Sprechen oder Sprechenlernen (auch Lesen und Schreiben) Schwierigkeiten hatte oder hat? Nein ☐ Ja ☐ (bitte kreuzen Sie an):

Person	Lispeln	Stottern	andere Störungen der Aussprache	verspäteter Sprechbeginn	LRS	andere Auffälligkeit (bitte nennen):
Leibliche Mutter	☐	☐	☐	☐	☐	☐
Leiblicher Vater	☐	☐	☐	☐	☐	☐
Oma (mütterl.)	☐	☐	☐	☐	☐	☐
Opa (mütterl.)	☐	☐	☐	☐	☐	☐
Oma (väterl.)	☐	☐	☐	☐	☐	☐
Opa (väterl.)	☐	☐	☐	☐	☐	☐
1. Geschwister	☐	☐	☐	☐	☐	☐
2. Geschwister	☐	☐	☐	☐	☐	☐
__________	☐	☐	☐	☐	☐	☐

13. Listen Sie **alle** Sprachen auf, die die/der Sorgeberechtigte 1 beherrscht und kreuzen Sie das entsprechende Niveau an. Außerdem schreiben Sie bitte in Klammern das Alter in Jahren, ab dem er/sie mit dieser Sprache regelmäßig in Kontakt gekommen ist: z.B. Deutsch (0 – seit der Geburt), Englisch (12).

Sprachkenntnisse (1 mind. und 10 max.)	1	2	3	4	5	6	7	8	9	10
Deutsch	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Italienisch	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
______________	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Welche der o.g. Sprachen spricht er/sie mit dem Kind? ______________________

14. Listen Sie **alle** Sprachen auf, die die/der Sorgeberechtigte 2 beherrscht und kreuzen Sie das entsprechende Niveau an. Außerdem schreiben Sie bitte in Klammern das Alter in Jahren, ab dem er/sie mit dieser Sprache regelmäßig in Kontakt gekommen ist: z.B. Deutsch (0 – seit der Geburt), Englisch (12).

Sprachkenntnisse (1 mind. und 10 max.)	1	2	3	4	5	6	7	8	9	10
Deutsch	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Italienisch	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
______________	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Welche der o.g. Sprachen spricht sie/er mit dem Kind? ______________________

15. Wohnen andere Personen mit im Haus? Nein ☐ Ja ☐
Wenn ja, nennen Sie deren Beziehung zum Kind (Bruder, Oma, neuer Lebensgefährte) und die Sprache/n, die mit dem Kind gesprochen wird. Im Fall von Geschwistern, schreiben Sie in Klammern wie alt sie sind.

Beziehung zum Kind	Sprache/n mit dem Kind gesprochen

16. Bitte kreuzen Sie den höchsten Bildungsabschluss der Sorgeberechtigten an:

Sorgeberechtigte/r 1	Sorgeberechtigte/r 2
1.____Grundschule	1.___Grundschule
2.____Mittlere Reife (10 Klasse) +/- Berufsausbild	2.___Mittlere Reife (10 Klasse) +/- Berufsausbild
3.____Abitur +/- Berufsausbild	3.___ Abitur +/- Berufsausbild
4.____Abgeschlossenes Studium	4.___ Abgeschlossenes Studium

Teil B – Sprachliche Erfahrung des Kindes

17. Bitte geben Sie an, wie viele Jahre und Monate das Kind in jeder Sprachumgebung verbracht hat:

	Jahre	Monate
Deutschland		
Italien		
Andere:		

18. Bitte listen Sie die Sprachen auf, die das Kind spricht. Beginnen Sie mit der Sprache, die **es am besten kann**, **bis zur Sprache, die es kaum/wenig kann**. Falls zwei Sprachen das gleiche Niveau haben, dann schreiben Sie sie in das gleiche Kästchen:

1.	2.	3.

19. Bitte listen Sie die Sprachen auf, die das Kind spricht. Beginnen Sie mit der Sprache, die **es am frühesten erworben hat, bis zur Sprache, die es am spätesten erworben hat**. Falls zwei Sprachen gleichzeitig erworben wurden, dann schreiben Sie sie in das gleiche Kästchen:

1.	2.	3.

20. Bitte schreiben Sie das Alter auf, ab dem das Kind mit der o.g. Sprachen regelmäßig in Kontakt gekommen ist: (z.B. Deutsch (0 – seit der Geburt), Italienisch (2)).

1.	2.	3.

21. Bitte schätzen Sie ein, wie gut Ihr Kind die folgenden Sprachen versteht?

	schlecht	befriedigend	gut	ausgezeichnet
Deutsch	☐	☐	☐	☐
Italienisch	☐	☐	☐	☐
Andere: ________________	☐	☐	☐	☐

22. Bitte schätzen Sie ein, wie gut Ihr Kind die folgenden Sprachen spricht?

	schlecht	befriedigend	gut	ausgezeichnet
Deutsch	☐	☐	☐	☐
Italienisch	☐	☐	☐	☐
Andere: ________________	☐	☐	☐	☐

23. Ist Ihr Kind in einem monolingualen oder mehrsprachigen Kindergarten eingeschrieben? ______________
Welche Sprache/n wird/werden dort gesprochen? ____

24. Nimmt Ihr Kind an zusätzlichen Sprachunterrichten teil? ja ☐ nein ☐
Wenn ja, für welche Sprache und wie viele Stunden in der Woche? ______________________________

25. Schätzen Sie bitte ein, wieviel Prozent der Zeit Ihr Kind jede Sprache **seit seiner Geburt gehört hat**. Sie können sich vorstellen, dass man ein Tonaufnahmegerät neben das Ohr Ihres Kindes gestellt hat. (Die Anteile sollten sich auf 100% in jeder Spalte summieren, z.B. „ab dem 4. Lebensjahr" Deutsch 70% Italienisch 30%).

	von der Geburt bis zum 2. Lebensjahr	vom 2. bis zum 4. Lebensjahr	ab dem 4. Lebensjahr
Deutsch			
Italienisch			
Andere: ________			

26. Bitte schätzen Sie ein, wie viele **Stunden pro Woche (Mo-So)** Ihr Kind mit den folgenden Personen/an den folgenden Orten/mit den folgenden Aktivitäten verbringt. Bitte denken Sie daran, dass die Schlafzeit nicht berücksichtigt werden soll. Falls etwas nicht zutrifft, streichen Sie bitte die Spalte durch.

	Sorgebe-rechtigte/r 1	Sorgebe-rechtigte/r 2	Sorgebe-rechtigte (zusammen)	Geschwis-ter	Im Kinder-garten	Sonstige Betreuer/innen	(Vor) lesen	TV/PC Spiele/Musik/ Internet	Sport/ Sonstiges
Stunden pro Woche									

27. Bitte schätzen Sie ein, wieviel Prozent der Zeit Ihr Kind für die folgenden **Aktivitäten** jede Sprache **benutzt** (die Anteile sollten sich auf 100% in jeder Spalte summieren, z.B. „(vor)lesen": Deutsch 60% Italienisch 30% Englisch 10%). Falls etwas nicht zutrifft, streichen Sie bitte die Spalte durch.

	(Vor)lesen	TV/PC Spiele/Musik/Internet	Sport/Sonstiges
Deutsch			
Italienisch			
Andere: ________			

A.2 Monolingual language background questionnaire

Fragebogen zum sozialen und sprachlichen Hintergrund

Liebe Eltern,

von den Kindern, die an unseren Studien teilnehmen, benötigen wir einige Informationen zu Entwicklung, Familie und Sprache. Alle Angaben werden vertraulich behandelt und nur in anonymisierter Form gespeichert. Bitte nehmen Sie sich ca. 20 Minuten Zeit, um den Fragebogen sorgfältig und aufmerksam auszufüllen. Falls Sie Fragen oder Wünsche haben, schreiben Sie sie bitte auf der letzten Seite (Teil C – Kommentare der Familie) auf.

1. Heutiges Datum: ____________________
Tag Monat Jahr

2. Ausgefüllt von: Sorgeberechtigte/r 1 ☐ Beziehung zum Kind bitte nennen: ____________________
Sorgeberechtigte/r 2 ☐ Beziehung zum Kind bitte nennen: ____________________
Andere ☐ (bitte ausführen) ____________________

Teil A – Hintergrund

Die folgenden Informationen beziehen sich auf Ihr Kind:

3. Geburtsdatum: ____________________
Tag Monat Jahr

4. Erfolgte die Geburt zum errechneten Termin? Ja ☐ Nein ☐
Wie viele Tage früher? ______________
Wie viele Tage später? ______________

5. Geschlecht: weiblich ☐ männlich ☐

6. Kita-/ Schuljahr: ____________________

7. Geburtsland: ____________________

8. Ist das Sehvermögen des Kindes beeinträchtigt?
Nein ☐ Ja ☐ (bitte Auffälligkeit nennen) ____________________

9. Hatte das Kind Krankheiten, die das Hörvermögen beeinträchtigen? (z.B. Mittelohrentzündungen)
Nein ☐ Ja ☐ (bitte Auffälligkeit nennen) ____________________
Wann das letzte Mal? Und wie häufig pro Jahr? ____________________

10. Gibt es diagnostizierte chronische Krankheiten?
Nein ☐ Ja ☐ (bitte Auffälligkeit nennen) ____________________

11. War das Kind jemals in einer sprachtherapeutischen Behandlung?
Nein ☐ Ja ☐ (bitte Auffälligkeit nennen) ____________________
Von wann bis wann erfolgte die Sprachtherapie? ____________________

28. Bitte schätzen Sie ein, wieviel Prozent der Zeit Ihr Kind jede Sprache von den folgenden Personen **hört**? (Die Anteile sollen sich auf 100% in jeder Spalte summieren, z.B. „Sorgeberechtigte/r 1": Deutsch 20% Italienisch 80%). Falls etwas nicht zutrifft, streichen Sie bitte die Spalte durch.

	Sorgeberechtigte/r 1	Sorgeberechtigte/r 2	Sorgeberechtigte untereinander
Deutsch			
Italienisch			
Andere: ________			

	Geschwister	Im Kindergarten	Sonstige Betreuer/innen
Deutsch			
Italienisch			
Andere: ________			

29. Bitte schätzen Sie ein, wieviel Prozent der Zeit Ihr Kind jede Sprache mit den folgenden Personen **spricht**? (Die Anteile sollen sich auf 100% in jeder Spalte summieren, z.B. „Sorgeberechtigte/r 1 Deutsch 40% Italienisch 60%). Falls etwas nicht zutrifft, streichen Sie bitte die Spalte durch.

	Sorgeberechtigte/r 1	Sorgeberechtigte/r 2
Deutsch		
Italienisch		
Andere: ________		

	Geschwister	Im Kindergarten	Sonstige Betreuer/innen
Deutsch			
Italienisch			
Andere: ________			

Teil C – Kommentare der Familie

____

____

Vielen Dank für Ihre wertvolle Unterstützung!

Die folgenden Informationen beziehen sich auf die Sorgeberechtigten:

12. Gibt es einen nahen Verwandten des Kindes, der mit dem Sprechen oder Sprechenlernen (auch Lesen und Schreiben) Schwierigkeiten hatte oder hat? Nein ☐ Ja ☐ (bitte kreuzen Sie an):

Person	Lispeln	Stottern	andere Störungen der Aussprache	verspäteter Sprechbeginn	LRS	andere Auffälligkeit (bitte nennen):
Leibliche Mutter	☐	☐	☐	☐	☐	☐
Leiblicher Vater	☐	☐	☐	☐	☐	☐
Oma (mütterl.)	☐	☐	☐	☐	☐	☐
Opa (mütterl.)	☐	☐	☐	☐	☐	☐
Oma (väterl.)	☐	☐	☐	☐	☐	☐
Opa (väterl.)	☐	☐	☐	☐	☐	☐
1. Geschwister	☐	☐	☐	☐	☐	☐
2. Geschwister	☐	☐	☐	☐	☐	☐
____________	☐	☐	☐	☐	☐	☐

13. Listen Sie **alle** Sprachen auf, die die/der Sorgeberechtigte 1 beherrscht und kreuzen Sie das entsprechende Niveau an. Außerdem schreiben Sie bitte in Klammern das Alter in Jahren, ab dem er/sie mit dieser Sprache regelmäßig in Kontakt gekommen ist: z.B. Deutsch (0 – seit der Geburt), Englisch (12).

Sprachkenntnisse (1 mind. und 10 max.)	1	2	3	4	5	6	7	8	9	10
Deutsch	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Italienisch	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
____________	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Welche der o.g. Sprachen spricht er/sie mit dem Kind? ____________________

14. Listen Sie **alle** Sprachen auf, die die/der Sorgeberechtigte 2 beherrscht und kreuzen Sie das entsprechende Niveau an. Außerdem schreiben Sie bitte in Klammern das Alter in Jahren, ab dem er/sie mit dieser Sprache regelmäßig in Kontakt gekommen ist: z.B. Deutsch (0 – seit der Geburt), Englisch (12).

Sprachkenntnisse (1 mind. und 10 max.)	1	2	3	4	5	6	7	8	9	10
Deutsch	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Italienisch	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
____________	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Welche der o.g. Sprachen spricht sie/er mit dem Kind? ____________________

15. Wohnen andere Personen mit im Haus? Nein ☐ Ja ☐
Wenn ja, nennen Sie deren Beziehung zum Kind (Bruder, Oma, neuer Lebensgefährte) und die Sprache/n, die mit dem Kind gesprochen wird. Im Fall von Geschwistern, schreiben Sie in Klammern wie alt sie sind.

Beziehung zum Kind	Sprache/n mit dem Kind gesprochen

16. Bitte kreuzen Sie den höchsten Bildungsabschluss der Sorgeberechtigten an:

Sorgeberechtigte/r 1	Sorgeberechtigte/r 2
1.____Grundschule	1.___Grundschule
2.____Mittlere Reife (10 Klasse) +/- Berufsausbild	2.___Mittlere Reife (10 Klasse) +/- Berufsausbild
3.____Abitur +/- Berufsausbild	3.___ Abitur +/- Berufsausbild
4.____Abgeschlossenes Studium	4.___ Abgeschlossenes Studium
5.____Promotionsabschluss	5.___ Promotionsabschluss

Teil C – Kommentare der Familie

____

____

Vielen Dank für Ihre wertvolle Unterstützung!

B OVS: material

The two versions of the OVS experiment.

OVS Version 1

Date: Code:

Date of birth: Sex:

pos	item	sentence	targ	com
1	10	Anna_Intro	-	
2	101	Schau mal, ein Schuh. Welche Farbe hat der Rahmen?	gelb	
3	103	Schau mal, ein Auge. Welche Farbe hat der Rahmen?	blau	
4	105	Schau mal, die Katze singt. Welche Farbe hat der Rahmen?	blau	
5	107	Die Schildkröte schaukelt. Welche Farbe hat der Rahmen?	blau	
6	109	Die Maus spielt Gitarre	gelb	
7	111	Die Schlange liest ein Buch	blau	
8	113	Anna_los geht's	-	
9	11111a	Der Falke zieht den Hund	gelb	
10	14881c	Den Elefanten fotografiert der Wolf	blau	
11	1145281d	Das Nashorn malt der Pandabär	blau	
12	1162301b	Das Kamel bespritzt den Waschbären	blau	
13	13661c	Den Eisbären fesselt der Delfin	gelb	
14	1137261d	Das Krokodil impft der Leopard	gelb	
15	12331a	Der Löwe pikst den Dino	gelb	
16	1154281b	Das Pony grüßt den Raben	blau	
17	114	Anna_weiter so	-	
18	1103191b	Das Küken pikst den Krebs	gelb	
19	1116221d	Das Seepferdchen fesselt der Papagei	gelb	
20	174131a	Der Pinguin grüßt den Bullen	gelb	
21	1128241d	Das Lama fotografiert der Fuchs	blau	
22	157101c	Den Igel impft der Hase	blau	
23	191171b	Das Huhn zieht den Hamster	gelb	
24	165121c	Den Hirsch malt der Koalabär	blau	
25	182151a	Der Fisch bespritzt den Affen	gelb	
26	115	Anna_die Hälfte	-	
27	12441c	Den Löwen grüßt der Dino	gelb	
28	14771a	Der Elefant im	blau	
29	1138251b	Das Krokodil fotografiert den Leoparden	blau	
30	13551a	Der Eisbär malt den Delfin	gelb	
31	1153291d	Das Pony pikst der Rabe	blau	
32	1146271b	Das Nashorn fesselt den Pandabären	gelb	
33	11221c	Den Falken bespritzt der Hund	blau	
34	1161311d	Das Kamel zieht der Waschbär	gelb	
35	116	Anna_bald fertig	-	
36	192181d	Das Huhn bespritzt der Hamster	blau	
37	181161c	Den Fisch zieht der Affe	blau	
38	1115211b	Das Seepferdchen malt den Papagei	gelb	
39	166111a	Der Hirsch fesselt den Koalabären	gelb	
40	1127231b	Das Lama impft den Fuschs	blau	
41	173141c	Den Pinguin pikst der Bulle	blau	
42	1104201d	Das Küken grüßt der Krebs	blau	
43	15891a	Der Igel fotografiert den Hasen	gelb	
44	117	Anna_Applaus	-	

OVS Version 2

Date: Code:

Date of birth: Sex:

pos	item	sentence	targ	com
1	10	Anna_Intro	-	
2	101	Schau mal, ein Schuh. Welche Farbe hat der Rahmen?	gelb	
3	103	Schau mal, ein Auge. Welche Farbe hat der Rahmen?	blau	
4	105	Schau mal, die Katze singt. Welche Farbe hat der Rahmen?	blau	
5	107	Die Schildkröte schaukelt. Welche Farbe hat der Rahmen?	blau	
6	109	Die Maus spielt Gitarre	gelb	
7	111	Die Schlange liest ein Buch	blau	
8	113	Anna_los geht's	-	
9	1162312	Das Kamel bespritzt der Waschbär	gelb	
10	1145272	Das Nashorn malt den Pandabären	gelb	
11	181152a	Der Fisch zieht den Affen	gelb	
12	1137252	Das Krokodil impft den Leoparden	blau	
13	166122a	Den Hirsch fesselt der Koalabär	blau	
14	1154292	Das Pony grüßt der Rabe	blau	
15	158102c	Den Igel fotografiert der Hase	gelb	
16	173132a	Der Pinguin pikst den Bullen	gelb	
17	114	Anna_weiter so	-	
18	191182d	Das Huhn zieht der Hamster	gelb	
19	13562c	Den Eisbären malt der Delfin	blau	
20	12432a	Der Löwe grüßt den Dino	blau	
21	1103202	Das Küken pikst der Krebs	blau	
22	11212a	Der Falke bespritzt den Hund	gelb	
23	1128232	Das Lama fotografiert den Fuchs	blau	
24	14782c	Den Elefanten impft der Wolf	blau	
25	1116212	Das Seepferdchen fesselt den Papagei	blau	
26	115	Anna_die Hälfte	-	

pos	item	sentence	targ	com
27	1161302	Das Kamel zieht den Waschbären	gelb	
28	15792a	Der Igel impft den Hasen	blau	
29	1138262	Das Krokodil fotografiert der Leopard	blau	
30	182162c	Den Fisch bespritzt der Affe	gelb	
31	1153282	Das Pony pikst den Raben	gelb	
32	1146282	Das Nashorn fesselt der Pandabär	gelb	
33	174142c	Den Pinguin grüßt der Bulle	blau	
34	165112a	Der Hirsch malt den Koalabären	gelb	
35	116	Anna_bald fertig	-	
36	1115222	Das Seepferdchen malt der Papagei	gelb	
37	13652a	Der Eisbär fesselt den Delfin	blau	
38	192172d	Das Huhn bespritzt den Hamster	gelb	
39	1127242	Das Lama impft der Fuschs	blau	
40	11122c	Den Falken zieht der Hund	blau	
41	1104192	Das Küken grüßt den Krebs	gelb	
42	14872a	Der Elefant fotografiert den Wolf	gelb	
43	12342c	Den Löwen pikst der Dino	blau	
44	117	Anna_Applaus	-	

C OVS: materials

C.1 Output statistical model Table 7

Treatment contrast for language group and condition. Monolinguals and OVS early in intercept. Random intercepts: participants and items. Random slope: individual of condition effect for each participant. Correct responses as dependent variable.

```
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) ['glmerMod']
Family: binomial  ( logit )
Formula: acc ~ 1 + lang_group * cond + (1 + cond|id) + (1|item)
Data: d
Control: glmerControl(calc.derivs = FALSE)

     AIC      BIC   logLik deviance df.resid
  1419.8   1492.5   -696.9   1393.8     1971

Scaled residuals:
    Min      1Q  Median      3Q     Max
-7.7293 -0.3383  0.1551  0.2566  5.5755

Random effects:
Groups Name        Variance Std.Dev. Corr
id     (Intercept) 1.81970  1.3490
        OVSlate    0.67641  0.8224 -0.39
        SVO        1.63464  1.2785 -0.75  -0.31
item   (Intercept) 0.09499  0.3082
Number of obs: 1984, groups:  id, 62; item, 64

Fixed effects:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)        0.8210     0.2948   2.785  0.00535 **
bili              -1.2620     0.4045  -3.120  0.00181 **
OVSlate           -1.7051     0.2817  -6.054 1.42e-09 ***
SVO                2.6005     0.3742   6.949 3.68e-12 ***
bili:OVSlate      -0.3964     0.4067  -0.975  0.32966
bili:SVO           1.4402     0.5224   2.757  0.00584 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.2 Output statistical model Table 8

Treatment contrast for language group and condition. Bilinguals and OVS late in intercept. Random intercepts: participants and items. Random slope: individual of condition effect for each participant. Correct responses as dependent variable.

```
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) ['glmerMod']
Family: binomial  ( logit )
Formula: acc ~ 1 + lang_group * cond + (1 + cond|id) + (1|item)
Data: d
Control: glmerControl(calc.derivs = FALSE)

     AIC      BIC   logLik deviance df.resid
  1419.8   1492.5   -696.9   1393.8     1971

Scaled residuals:
    Min      1Q  Median      3Q     Max
-7.7293 -0.3383  0.1551  0.2566  5.5755

Random effects:
Groups Name        Variance Std.Dev. Corr
id     (Intercept) 1.62834  1.2761
        OVSearly   0.67096  0.8191 -0.23
        SVO        2.95226  1.7182 -0.85  0.71
item   (Intercept) 0.09503  0.3083
Number of obs: 1984, groups:  id, 62; item, 64

Fixed effects:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)       -2.5400     0.3338  -7.609 2.75e-14 ***
mono               1.6562     0.4256   3.892 9.95e-05 ***
OVSearly           2.0995     0.3310   6.343 2.26e-10 ***
SVO                6.1403     0.4731  12.979  < 2e-16 ***
mono:OVSearly     -0.3948     0.4061  -0.972  0.33092
mono:SVO          -1.8347     0.6252  -2.934  0.00334 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.3 Output statistical model Table 9

Treatment contrast for language group and condition. Monolinguals and OVS early in intercept. Random intercepts: participants and items. Model output for first time window NP1. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "NP1")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 4789.4

Scaled residuals: 
     Min       1Q   Median       3Q      Max 
-1.17336 -1.08484  0.03562  1.08703  1.25676 

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept) 0.0000   0.000   
 id       (Intercept) 0.0000   0.000   
 Residual             0.7849   0.886   
Number of obs: 1838, groups:  item, 64; id, 62

Fixed effects:
                  Estimate Std. Error         df t value Pr(>|t|)
(Intercept)        0.03692    0.05842 1832.00000   0.632    0.527
bili              -0.03105    0.08271 1832.00000  -0.375    0.707
OVSlate           -0.07579    0.08167 1832.00000  -0.928    0.354
SVO               -0.01562    0.07155 1832.00000  -0.218    0.827
bili:OVSlate      -0.04354    0.11643 1832.00000  -0.374    0.708
bili:SVO           0.04930    0.10143 1832.00000   0.486    0.627
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.4 Output statistical model Table 9

Treatment contrast for language group and condition. Monolinguals and OVS early in intercept. Random intercepts: participants and items. Model output for the time window V. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "V")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 4847

Scaled residuals: 
    Min      1Q  Median      3Q     Max 
-1.4385 -1.0130  0.2280  0.8968  1.3577 

Random effects:
 Groups   Name        Variance  Std.Dev. 
 item     (Intercept) 7.533e-03 8.679e-02
 id       (Intercept) 6.338e-17 7.961e-09
 Residual             8.250e-01 9.083e-01
Number of obs: 1820, groups:  item, 64; id, 62

Fixed effects:
               Estimate Std. Error         df t value Pr(>|t|)
(Intercept)  -4.417e-02  6.413e-02  1.874e+02  -0.689 0.491809
bili         -9.406e-03  8.557e-02  1.761e+03  -0.110 0.912482
OVSlate      -6.494e-02  8.982e-02  1.818e+02  -0.723 0.470580
SVO           2.624e-01  7.838e-02  1.864e+02   3.348 0.000984***
bili:OVSlate -6.562e-02  1.203e-01  1.762e+03  -0.545 0.585553
bili:SVO     -2.120e-02  1.047e-01  1.763e+03  -0.202 0.839602
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.5 Output statistical model Table 9

Treatment contrast for language group and condition. Monolinguals and OVS early in intercept. Random intercepts: participants and items. Model output for the time window NP2. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "NP2")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 4574.3

Scaled residuals:
     Min       1Q   Median     3Q     Max
 -1.8291  -0.8201  0.3162  0.7281  1.7957

Random effects:
 Groups   Name         Variance Std.Dev.
 item     (Intercept) 0.0103901 0.10193
 id       (Intercept) 0.0007004 0.02646
 Residual             0.6909261 0.83122
Number of obs: 1836, groups:  item, 64; id, 62

Fixed effects:
                Estimate Std. Error         df t value Pr(>|t|)
(Intercept)      0.05980    0.06082  157.12051   0.983   0.3270
bili            -0.13882    0.07883  698.15572  -1.761   0.0787.
OVSlate         -0.40908    0.08512  167.82274  -4.806 3.40e-06***
SVO              0.37166    0.07410  170.56710   5.016 1.32e-06***
bili:OVSlate     0.09510    0.11040 1729.99749   0.861   0.3891
bili:SVO         0.09970    0.09567 1730.71459   1.042   0.2975
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.6 Output statistical model Table 9

Treatment contrast for language group and condition. Monolinguals and OVS early in intercept. Random intercepts: participants and items. Model output for time window Silence 1. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "Sil1")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 4389.9

Scaled residuals:
     Min       1Q   Median     3Q     Max
 -2.2046  -0.7744   0.4002  0.6120 2.0900

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept) 0.01085  0.1042
 id       (Intercept) 0.01036  0.1018
 Residual             0.60323  0.7767
Number of obs: 1852, groups:  item, 64; id, 62

Fixed effects:
               Estimate Std. Error         df t value Pr(>|t|)
(Intercept)     0.20693    0.06069  167.35128   3.410 0.000815***
bili           -0.28981    0.07736  472.49058  -3.746 0.000202***
OVSlate        -0.54439    0.08096  161.32327  -6.724 2.90e-10***
SVO             0.36561    0.07052  164.42851   5.184 6.29e-07***
bili:OVSlate    0.14304    0.10254 1740.77281   1.395 0.163205
bili:SVO        0.29852    0.08892 1740.78297   3.357 0.000804***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.7 Output statistical model Table 9

Treatment contrast for language group and condition. Monolinguals and OVS early in intercept. Random intercepts: participants and items. Model output for time window Silence 2. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "Sil2")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 4440.8

Scaled residuals:
     Min       1Q   Median     3Q      Max
 -2.2774  -0.9201   0.3892  0.6426  1.9178

Random effects:
 Groups   Name         Variance Std.Dev.
 item     (Intercept)  0.004198 0.06479
 id       (Intercept)  0.020027 0.14152
 Residual              0.632220 0.79512
Number of obs: 1837, groups:  item, 64; id, 62

Fixed effects:
                Estimate Std. Error         df t value Pr(>|t|)
(Intercept)      0.25466    0.06084  181.41343   4.186 4.42e-05***
bili            -0.36580    0.08291  364.63497  -4.412 1.35e-05***
OVSlate         -0.40771    0.07785  197.85883  -5.238 4.15e-07***
SVO              0.33671    0.06751  198.04798   4.987 1.34e-06***
bili:OVSlate     0.18463    0.10528 1723.22075   1.754 0.079659.
bili:SVO         0.33695    0.09123 1724.65253   3.693 0.000228***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.8 Output statistical model Table 9

Treatment contrast for language group and condition. Monolinguals and OVS early in intercept. Random intercepts: participants and items. Model output for time window Silence 3. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "Sil3")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 4529.7

Scaled residuals:
     Min       1Q  Median     3Q     Max
 -1.7454  -1.0679  0.5068  0.7999  1.4570

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept) 0.002952 0.05433
 id       (Intercept) 0.007627 0.08733
 Residual             0.740757 0.86067
Number of obs: 1769, groups:  item, 64; id, 62

Fixed effects:
                Estimate Std. Error         df t value Pr(>|t|)
(Intercept)      0.19519    0.06035  163.12336   3.234 0.001476**
bili            -0.20877    0.08439  564.85628  -2.474 0.013665*
OVSlate         -0.21754    0.08292  191.39556  -2.623 0.009407**
SVO              0.24091    0.07207  193.88646   3.343 0.000995***
bili:OVSlate     0.05110    0.11536 1659.89480   0.443 0.657855
bili:SVO         0.09954    0.10006 1663.89847   0.995 0.319958
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.9 Output statistical model Table 10

Treatment contrast for language group and condition. Bilinguals and OVS late in intercept. Random intercepts: participants and items. Model output for first time window NP1. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "NP1")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 4789.4

Scaled residuals:
     Min       1Q   Median       3Q      Max
-1.7336  -1.08484  0.03562  1.08703  1.25676

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept) 0.0000   0.000
 id       (Intercept) 0.0000   0.000
 Residual             0.7849   0.886
Number of obs: 1838, groups:  item, 64; id, 62

Fixed effects:
                Estimate Std. Error         df t value Pr(>|t|)
(Intercept)     -0.11345    0.05880 1832.00000  -1.929   0.0538.
mono             0.07459    0.08194 1832.00000   0.910   0.3628
OVSearly         0.11933    0.08298 1832.00000   1.438   0.1506
SVO              0.15301    0.07210 1832.00000   2.122   0.0340*
mono:OVSearly   -0.04354    0.11643 1832.00000  -0.374   0.7085
mono:SVO        -0.09284    0.10081 1832.00000  -0.921   0.3572
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.10 Output statistical model Table 10

Treatment contrast for language group and condition. Bilinguals and OVS late in intercept. Random intercepts: participants and items. Model output for time window V. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "V")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 4847

Scaled residuals:
     Min       1Q   Median       3Q     Max
-1.4385   -1.0130   0.2280   0.8968  1.3577

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept) 0.007533 0.08679
 id       (Intercept) 0.0000   0.000
 Residual             0.825013 0.90830
Number of obs: 1820, groups:  item, 64; id, 62

Fixed effects:
                Estimate Std. Error         df t value Pr(>|t|)
(Intercept)     -0.18414    0.06434  192.71778  -2.862  0.00468**
mono             0.07503    0.08458 1762.49182   0.887  0.37517
OVSearly         0.13056    0.09108  193.43028   1.433  0.15335
SVO              0.37180    0.07888  193.37978   4.713 4.65e-06***
mono:OVSearly   -0.06562    0.12031 1761.94309  -0.545  0.58555
mono:SVO        -0.04442    0.10390 1763.36507  -0.428  0.66905
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.11 Output statistical model Table 10

Treatment contrast for language group and condition. Bilinguals and OVS late in intercept. Random intercepts: participants and items. Model output for time window NP2. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "NP2")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 4574.3

Scaled residuals:
     Min      1Q  Median      3Q     Max
 -1.8291 -0.8201  0.3162  0.7281  1.7957

Random effects:
 Groups   Name        Variance  Std.Dev.
 item     (Intercept) 0.0103901 0.10193
 id       (Intercept) 0.0007004 0.02646
 Residual             0.6909261 0.83122
Number of obs: 1836, groups:  item, 64; id, 62

Fixed effects:
                Estimate Std. Error        df t value Pr(>|t|)
(Intercept)   -3.930e-01  6.143e-02  1.653e+02  -6.398 1.55e-09***
mono           4.372e-02  7.788e-02  6.802e+02   0.561 0.574766
OVSearly       3.140e-01  8.685e-02  1.830e+02   3.615 0.000388***
SVO            7.853e-01  7.464e-02  1.777e+02  10.521  < 2e-16***
mono:OVSearly  9.510e-02  1.104e-01  1.730e+03   0.861 0.389113
mono:SVO      -4.596e-03  9.489e-02  1.724e+03  -0.048 0.961375
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.12 Output statistical model Table 10

Treatment contrast for language group and condition. Bilinguals and OVS late in intercept. Random intercepts: participants and items. Model output for time window Silence 1. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "Sil1")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 4389.9

Scaled residuals:
     Min      1Q  Median      3Q     Max
 -2.2046 -0.7744  0.4002  0.6120  2.0900

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept) 0.01085  0.1042
 id       (Intercept) 0.01036  0.1018
 Residual             0.60323  0.7767
Number of obs: 1852, groups:  item, 64; id, 62

Fixed effects:
                Estimate Std. Error        df t value Pr(>|t|)
(Intercept)     -0.48422    0.06072  169.91752  -7.975 2.14e-13***
mono             0.14676    0.07663  460.38292   1.915   0.0561.
OVSearly         0.40134    0.08173  168.84986   4.911 2.13e-06***
SVO              1.06548    0.07072  168.36312  15.066  < 2e-16***
mono:OVSearly    0.14304    0.10254 1740.77281   1.395   0.1632
mono:SVO        -0.15548    0.08831 1742.58376  -1.761   0.0785.
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.13 Output statistical model Table 10

Treatment contrast for language group and condition. Bilinguals and OVS late in intercept. Random intercepts: participants and items. Model output for time window Silence 2. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "Sil2")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 4440.8

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.2774 -0.9201  0.3892  0.6426  1.9178

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept) 0.004198 0.06479
 id       (Intercept) 0.020027 0.14152
 Residual             0.632220 0.79512
Number of obs: 1837, groups:  item, 64; id, 62

Fixed effects:
                Estimate Std. Error        df t value Pr(>|t|)
(Intercept)     -0.33423    0.06061  180.92964  -5.514  1.2e-07***
mono             0.18117    0.08246  357.83877   2.197  0.02865*
OVSearly         0.22308    0.07795  201.29188   2.862  0.00466**
SVO              0.89674    0.06731  199.06463  13.322  < 2e-16***
mono:OVSearly    0.18463    0.10528 1723.22075   1.754  0.07966.
mono:SVO        -0.15232    0.09083 1721.39203  -1.677  0.09372.
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.14 Output statistical model Table 10

Treatment contrast for language group and condition. Bilinguals and OVS late in intercept. Random intercepts: participants and items. Model output for time window Silence 3. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "Sil3")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 4529.7

Scaled residuals:
    Min      1Q  Median      3Q     Max
 -1.7454  -1.0679  0.5068  0.7999  1.4570

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept) 0.002952 0.05433
 id       (Intercept) 0.007627 0.08733
 Residual             0.740757 0.86067
Number of obs: 1769, groups:  item, 64; id, 62

Fixed effects:
                Estimate Std. Error        df t value Pr(>|t|)
(Intercept)     -0.18001    0.06180  179.68949  -2.913  0.00404**
mono             0.15767    0.08470  569.86568   1.861  0.06320.
OVSearly         0.16644    0.08469  209.48597   1.965  0.05071.
SVO              0.50690    0.07308  205.89326   6.936 5.11e-11***
mono:OVSearly    0.05110    0.11536 1659.89481   0.443  0.65785
mono:SVO        -0.04845    0.10030 1652.94666  -0.483  0.62916
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.15 Output statistical model chance level monolingual-OVS late
Treatment contrast for language group and condition. Monolinguals and OVS late in intercept. Random intercepts: participants and items.
Proportion of looks to target as dependent variable.

Time window	Fixed effect	Estimate	SE	df	*t*-value	*p*-value
NP1	Intercept (Mono, OVS late)	-0.039	0.057	1832	-0.681	=.496
	OVS late (Bili)	-0.075	0.082	1832	-0.910	=.363
	OVS early (Mono)	0.076	0.082	1832	0.928	=.354
	SVO (Mono)	0.016	0.070	1832	0.854	=.393
	Bili x OVS early	0.044	0.116	1832	0.374	=.708
	Bili x SVO	0.093	0.101	1832	0.921	=.357
V	Intercept (Mono, OVS late)	-0.109	0.063	176.2	-1.735	=.085‡
	OVS late (Bili)	-0.075	0.085	1762	-0.887	=.375
	OVS early (Mono)	0.065	0.090	181.8	0.723	=.471
	SVO (Mono)	0.327	0.078	178.9	4.231	<.001***
	Bili x OVS early	0.066	0.120	1762	0.545	=.586
	Bili x SVO	0.044	0.104	1763	0.428	=.669
NP2	Intercept (Mono, OVS late)	-0.349	0.060	150.2	-5.829	<.001***
	OVS late (Bili)	-0.044	0.079	680.2	-0.561	=.575
	OVS early (Mono)	0.409	0.085	167.8	4.806	<.001***
	SVO (Mono)	0.781	0.073	165.7	10.642	<.001***
	Bili x OVS early	-0.095	0.110	1730	-0.861	=.389
	Bili x SVO	0.005	0.095	1724	0.048	=.961
Sil1	Intercept (Mono, OVS late)	-0.337	0.060	156.7	-5.670	<.001***
	OVS late (Bili)	-0.147	0.077	460.4	-1.915	=.056‡
	OVS early (Mono)	0.544	0.081	161.3	6.724	<.001***
	SVO (Mono)	0.910	0.070	156.9	13.09	<.001***
	Bili x OVS early	-0.143	0.103	1741	-1.395	=.163
	Bili x SVO	0.155	0.089	1743	1.761	=.079‡
Sil2	Intercept (Mono, OVS late)	-0.153	0.060	177.9	-2.533	=.012*
	OVS late (Bili)	-0.181	0.082	357.8	-2.197	=.029*
	OVS early (Mono)	0.408	0.078	197.9	5.238	<.001***
	SVO (Mono)	0.744	0.067	195.6	11.087	<.001***
	Bili x OVS early	-0.185	0.105	1723	-1.754	=.080‡
	Bili x SVO	0.152	0.091	1721	1.677	=.094‡
Sil3	Intercept (Mono, OVS late)	-0.022	0.060	171.3	-0.366	=.715
	OVS late (Bili)	-0.158	0.085	569.9	-1.861	=.063‡
	OVS early (Mono)	0.218	0.083	191.4	2.623	=.009**
	SVO (Mono)	0.458	0.073	200.6	6.311	<.001***

Time window	Fixed effect	Estimate	SE	df	*t*-value	*p*-value
	Bili x OVS early	-0.051	0.115	1660	-0.443	=.658
	Bili x SVO	0.048	0.100	1653	0.483	=.629

C.16 Output statistical model chance level bilingual-OVS early
Treatment contrast for language group and condition. Bilinguals and OVS early in intercept. Random intercepts: participants and items.
Proportion of looks to target as dependent variable.

Time window	Fixed effect	Estimate	SE	df	*t*-value	*p*-value
NP1	Intercept (Bili, OVS early)	0.006	0.059	1832	0.100	=.920
	OVS early (Mono)	0.031	0.083	1832	0.375	=.707
	OVS late (Bili)	0.119	0.083	1832	-1.438	=.151
	SVO (Bili)	0.034	0.072	1832	0.469	=.639
	Mono x OVS late	0.044	0.116	1832	0.374	=.709
	Mono x SVO	-0.049	0.101	1832	-0.486	=.627
V	Intercept (Bili, OVS early)	-0.054	0.065	194.1	-0.831	=.407
	OVS early (Mono)	0.009	0.085	1761	0.110	=.912
	OVS late (Bili)	-0.131	0.091	193.4	-1.433	=.153
	SVO (Bili)	0.241	0.079	194.3	3.054	=.003**
	Mono x OVS late	-0.066	0.120	1762	0.545	=.586
	Mono x SVO	0.021	0.105	1763	0.202	=.840
NP2	Intercept (Bili, OVS early)	-0.079	0.062	168.7	-1.279	0.203
	OVS early (Mono)	0.139	0.079	698.2	1.761	=.079
	OVS late (Bili)	-0.314	0.087	183	-3.615	<.001***
	SVO (Bili)	0.471	0.075	180.4	6.291	<.001***

Time window	**Fixed effect**	**Estimate**	**SE**	**df**	***t*-value**	***p*-value**
	Mono x OVS late	-0.095	0.110	1730	-0.861	=.389
	Mono x SVO	-0.099	0.096	1731	-1.042	=.297
Sil1	Intercept (Bili, OVS early)	-0.083	0.061	167.6	-1.369	=.173
	OVS early (Mono)	0.290	0.077	472.5	3.746	<.001***
	OVS late (Bili)	-0.410	0.082	168.8	-4.911	<.001***
	SVO (Bili)	0.664	0.071	166.6	9.414	<.001***
	Mono x OVS late	-0.143	0.103	1741	-1.395	=.163
	Mono x SVO	-0.299	0.089	1741	-3.357	<.001***
Sil2	Intercept (Bili, OVS early)	-0.111	0.061	183.1	-1.827	=.069±
	OVS early (Mono)	0.366	0.083	364.6	4.412	<.001***
	OVS late (Bili)	-0.223	0.078	201.3	-2.862	=.005**
	SVO (Bili)	0.674	0.068	201	9.98	<.001***
	Mono x OVS late	-0.185	0.105	1723	-1.754	=.080‡
	Mono x SVO	-0.337	0.091	1725	-3.693	<.001***
Sil3	Intercept (Bili, OVS early)	-0.014	0.062	183.8	-0.219	=.827
	OVS early (Mono)	0.209	0.084	564.9	2.474	=.014*
	OVS late (Bili)	-0.167	0.085	209.5	-1.965	=.051‡
	SVO (Bili)	0.340	0.073	209.6	4.644	<.001***

Time window	**Fixed effect**	**Estimate**	**SE**	**df**	***t*-value**	***p*-value**
	Mono x OVS late	-0.051	0.115	1660	-0.443	=.658
	Mono x SVO	-0.100	0.100	1664	-0.995	=.320

C.17 Output statistical model Table 13

Repeated contrasts used for condition and language group. Treatment contrast used for accuracy: correct responses in intercept. Random intercepts: participants and items. Random slope: individual adjustment of condition effect for each participant. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + cond * lang_group * acc + (1+cond|id)+(1|item)
Data: d
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 6143.4

Scaled residuals:
    Min      1Q   Median      3Q     Max
-1.7861 -0.7832  -0.3240  0.8290  1.8415

Random effects:
 Groups   Name         Variance Std.Dev. Corr
 id       (Intercept) 0.0007834 0.02799
          OVSlt-OVSel 0.0067356 0.08207  0.11
 item     (Intercept) 0.0012617 0.03552
 Residual             0.1758672 0.41937
Number of obs: 5475, groups:  id, 62; item, 32

Fixed effects:
                Estimate Std. Error          df t value Pr(>|t|)
(Intercept)    6.452e-01  1.383e-02  1.434e+02  46.636   <2e-16***
OVSlt-OVSel   -4.414e-02  2.889e-02  1.542e+02  -1.528    0.129
bi-mo          3.728e-04  2.462e-02  3.263e+02   0.015    0.988
inc.res       -3.107e-01  1.448e-02  1.353e+03 -21.456   <2e-16***
OVSlt-OVSel:
bi-mo         -1.696e-03  5.196e-02  2.896e+02  -0.033    0.974
OVSlt-OVSel:
inc.res        1.276e-02  2.940e-02  1.787e+03   0.434    0.664
bi-mo:
inc.res       -7.237e-03  2.884e-02  1.374e+03  -0.251    0.802
OVSlt-OVSel:bi-mo:
inc.res        3.691e-02  5.852e-02  1.812e+03   0.631    0.528
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.18 Output statistical model Table 14

Repeated contrasts for condition. Monolinguals' data.

Model output for time window 1. Random intercepts: participants and items.

Pupil dilation data as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: pup ~ 1 + cond + (1|id) + (1|item)
   Data: subset(d2mono, time == "1")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 5375132

Scaled residuals:
     Min      1Q   Median       3Q       Max
-13.3410  -0.4570   0.0031   0.4818  12.5975

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept)  1540     39.24
 id       (Intercept)  1568     39.60
 Residual             20229    142.23
Number of obs: 421443, groups:  item, 64; id, 31

Fixed effects:
                Estimate  Std. Error        df t value Pr(>|t|)
(Intercept)       98.0973     8.8054   61.0368  11.141 2.42e-16
OVSel-SVO         -0.5913    12.0274   60.3825  -0.049    0.961
OVSlt-OVSel       10.9826    13.8877   60.3750   0.791    0.432
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.19 Output statistical model Table 14
Repeated contrasts for condition. Monolinguals' data.
Model output for time window 2. Random intercepts: participants and items.
Pupil dilation data as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: pup ~ 1 + cond + (1|id) + (1|item)
   Data: subset(d2mono, time == "2")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 5368111

Scaled residuals:
     Min      1Q   Median       3Q      Max
-12.4784  -0.4355   0.0338   0.5081  12.1658

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept)  1724     41.53
 id       (Intercept)  2426     49.25
 Residual             23617    153.68
Number of obs: 415842, groups:  item, 64; id, 31

Fixed effects:
                Estimate Std. Error       df t value Pr(>|t|)
(Intercept)    107.9652     10.4137  53.0883  10.368  2.3e-14 ***
OVSel-SVO        9.5221    12.7288  60.2362   0.748    0.457
OVSlt-OVSel     -0.4569    14.6975  60.2286  -0.031    0.975
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.20 Output statistical model Table 15
Repeated contrasts for condition. Bilinguals' data.
Model output for time window 1. Random intercepts: participants and items.
Pupil dilation data as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: pup ~ 1 + cond + (1|id) + (1|item)
   Data: subset(d2bili, time == "1")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 5919641

Scaled residuals:
     Min       1Q   Median       3Q      Max
-17.5261  -0.2352  -0.0032   0.2785   8.0243

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept)  3680     60.66
 id       (Intercept)  4538     67.36
 Residual             64964    254.88
Number of obs: 425238, groups:  item, 64; id, 31

Fixed effects:
               Estimate Std. Error     df t value Pr(>|t|)
(Intercept)       66.47      14.51  55.97   4.582 2.63e-05***
OVSel-SVO        -17.90      18.60  60.25  -0.963   0.3396
OVSlt-OVSel       44.79      21.48  60.25   2.085   0.0413 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.21 Output statistical model Table 15

Repeated contrasts for condition. Bilinguals' data.
Model output for time window 2. Random intercepts: participants and items.
Pupil dilation data as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: pup ~ 1 + cond + (1|id) + (1|item)
   Data: subset(d2bili, time == "2")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 5895370

Scaled residuals:
     Min       1Q   Median       3Q      Max
-16.2383  -0.2270   0.0006   0.2690   9.3945

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept)  5080     71.27
 id       (Intercept)  6505     80.65
 Residual             73927    271.90
Number of obs: 419597, groups:  item, 64; id, 31

Fixed effects:
                Estimate Std. Error       df t value Pr(>|t|)
(Intercept)       75.552     17.271   55.098   4.374 5.46e-05 ***
OVSel-SVO          4.755     21.847   60.248   0.218    0.828
OVSlt-OVSel       24.184     25.228   60.251   0.959    0.342
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.22 Output statistical model Table 16

Monolinguals' data. Repeated contrasts for condition. Treatment contrast used for accuracy: correct responses in intercept. Random intercepts: participants and items. Random slope: individual adjustment of condition effect for each participant. Pupil dilation as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + cond * acc + (1+cond|id)+(1|item)
Data: dmono
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 5392655

Scaled residuals:
     Min       1Q    Median       3Q      Max
-14.0236  -0.4512   0.0189   0.5085  13.3516

Random effects:
 Groups   Name        Variance   Std.Dev.   Corr
 id       (Intercept)   846.7     29.10
          OVSlt-OVSel  7290.7     85.39   0.38
 item     (Intercept)   3727.2    61.05
 Residual              18690.7  136.71
Number of obs: 425452, groups:  item, 32; id, 31

Fixed effects:
                Estimate Std. Error         df t value Pr(>|t|)
(Intercept)    1.057e+02  1.213e+01  4.230e+01   8.721 5.33e-11
...
OVSlt-OVSel   -4.278e+00  1.850e+01  5.149e+01  -0.231   0.8180
inc.res       -1.230e+00  5.688e-01  4.253e+05  -2.162   0.0306 *
OVSlt-OVSel:
inc.res        3.200e+01  1.137e+00  4.250e+05  28.128  < 2e-16
...
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.23 Output statistical model Table 17

Bilinguals' data. Repeated contrasts for condition. Treatment contrast used for correct responses in intercept. Random intercepts: participants and items. Random slope: individual adjustment of condition effect for each participant. Pupil dilation as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + cond * acc + (1+cond|id)+(1|item)
Data: dbili
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 5888992

Scaled residuals:
     Min      1Q    Median       3Q      Max
-14.6836  -0.2533  -0.0056   0.2465   7.4619

Random effects:
 Groups   Name        Variance   Std.Dev.   Corr
 id       (Intercept)    4700     68.56
          OVSlt-OVSel   28089    167.60   -0.57
 item     (Intercept)   11691    108.12
 Residual               71661    267.70
Number of obs: 420071, groups:  item, 32; id, 31

Fixed effects:
                Estimate Std. Error         df t value Pr(>|t|)
(Intercept)       66.014     22.925     49.981   2.880  0.00585 **
OVSlt-OVSel       79.031     38.725     56.675   2.041  0.04594 *
inc.res           18.697      1.398 419855.155  13.378  < 2e-16
OVSlt-OVSel:
inc.res          -68.381      2.795 419452.376 -24.465  < 2e-16
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.24 Table Degree of balanced bilingualism in 5-year-old group

Bold values indicate children that are considered to be balanced bilinguals.

ID	Ratio of German PPVT to Italian	Ratio of German TROG to Italian
1	1.48	1.99
2	1.38	1.20
3	1.48	**1.08**
4	1.68	**1.14**
5	1.64	1.16
6	3.10	1.89
7	1.17	**0.86**
8	1.61	**0.94**
9	2.28	**1.08**
10	1.50	1.22
11	1.68	**1.11**
12	1.27	0.84
13	2.36	1.24
14	3.39	1.70
15	1.13	**1.03**
16	3.33	**1.06**
17	2.84	1.19
18	2.13	1.35
19	2.73	1.30
20	2.40	1.17
21	1.56	1.16
22	2.14	1.34
23	2.24	**1.09**
24	2.15	**1.11**
25	2.00	0.52
26	3.00	1.30
27	1.82	**0.92**
28	1.42	**1.02**
29	2.42	**0.98**
30	1.26	**1.04**
31	1.38	**0.88**

C.25 Output statistical model Table 20

Repeated contrasts used for condition. Language group coded with treatment contrast: monolinguals in intercept. Random intercepts: participants and items. Random slope: individual adjustment of condition effect for each participant. Continuous predictors: flanker effect and global switch cost. Correct responses as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
 Family: binomial  ( logit )
Formula: acc~1+lang_group*cond*(Flanker+GSC)+(1+cond|id)+(1|item)
   Data: d
Control: glmerControl(calc.derivs = FALSE)

     AIC      BIC   logLik deviance df.resid
  1071.6   1150.0   -519.8   1039.6      976

Scaled residuals:
     Min      1Q    Median       3Q      Max
 -2.3213 -0.5589   -0.2579   0.5734   5.2273

Random effects:
 Groups   Name        Variance  Std.Dev.   Corr
 id       (Intercept) 1.16592   1.0798
          OVSlt-OVSel 0.59218   0.7695    -0.20
 item     (Intercept) 0.09327   0.3054
Number of obs: 992, groups:  id, 62; item, 32

Fixed effects:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)              -0.007749   0.231323  -0.034 0.973275
bi                       -1.449244   0.330879  -4.380 1.19e-05***
OVSlt-OVSel              -1.634981   0.279680  -5.846 5.04e-09***
Flanker                  -0.486752   0.746287  -0.652 0.514252
GSC                       1.304523   0.585055   2.230 0.025764*
bi:OVSlt-OVSel           -0.403491   0.407846  -0.989 0.322506
bi:Flanker                0.296328   1.000158   0.296 0.767016
bi:GSC                   -2.811222   0.851523  -3.301 0.000962***
OVSlt-OVSel:Flanker      -0.624566   0.873306  -0.715 0.474501
OVSlt-OVSel:GSC           0.562433   0.712000   0.790 0.429567
bi:OVSlt-OVSel:Flanker   -0.001056   1.224431  -0.001 0.999312
bi:OVSlt-OVSel:GSC       -0.065593   1.105675  -0.059 0.952694
```

C.26 Output statistical model Table 21

Repeated contrasts used for condition. Language group coded with treatment contrast: bilinguals in intercept. Random intercepts: participants and items. Random slope: individual adjustment of condition effect for each participant. Continuous predictors: flanker effect and global switch cost. Correct resp as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
 Family: binomial  ( logit )
Formula: acc~1+lang_group*cond*(Flanker+GSC)+(1+cond|id)+(1|item)
   Data: d
Control: glmerControl(calc.derivs = FALSE)

     AIC      BIC   logLik deviance df.resid
  1071.6   1150.0   -519.8   1039.6      976

Scaled residuals:
     Min      1Q    Median     3Q      Max
 -2.3159 -0.5595   -0.2572  0.5735  5.2286

Random effects:
 Groups   Name        Variance Std.Dev.   Corr
 id       (Intercept) 1.16568  1.0797
          OVSlt-OVSel 0.59781  0.7732    -0.20
 item     (Intercept) 0.09262  0.3043
Number of obs: 992, groups:  id, 62; item, 32

Fixed effects:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)              -1.456259   0.248601  -5.857 4.70e-09***
mo                        1.449030   0.330873   4.379 1.19e-05***
OVSlt-OVSel              -2.041872   0.334200  -6.110 9.98e-10***
Flanker                  -0.188298   0.666824  -0.282 0.777652
GSC                      -1.500049   0.618117  -2.427 0.015232*
mo:OVSlt-OVSel            0.409666   0.408385   1.003 0.315795
mo:Flanker               -0.313066   1.000252  -0.313 0.754291
mo:GSC                    2.808349   0.851359   3.299 0.000971***
OVSlt-OVSel:Flanker      -0.634204   0.863331  -0.735 0.462582
OVSlt-OVSel:GSC           0.506399   0.843809   0.600 0.548417
mo:OVSlt-OVSel:Flanker    0.007659   1.226425   0.006 0.995018
mo:OVSlt-OVSel:GSC        0.049308   1.106279   0.045 0.964449
```

C.27 Output statistical model Table 23

Repeated contrasts used for condition. Only the bilingual group is included. Random intercepts: participants and items. Random slope: individual adjustment of condition for each participant. Continuous predictors: language experience, number of older siblings and parents' education. Correct responses as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
 Family: binomial  ( logit )
 Formula:
 acc~1+cond*(lang_exp+L_edu+no_older_sib)+(1+cond|id)+(1|item)
 Data: df
Control: glmerControl(calc.derivs = FALSE)

     AIC      BIC   logLik deviance df.resid
   425.6    475.2   -200.8    401.7      452

Scaled residuals:
     Min      1Q    Median     3Q      Max
-2.2710  -0.4246   -0.2828 0.3331   4.4411

Random effects:
 Groups   Name        Variance  Std.Dev.   Corr
 item   (Intercept)    0.1210   0.3479
 id     (Intercept)    1.3952   1.1812
         OVSlt-OVSel  0.5437    0.7374   -1.00
Number of obs: 464, groups:  item, 32; id, 29

Fixed effects:
                           Estimate Std. Error z value Pr(>|z|)
(Intercept)               -1.55088    0.28136  -5.512 3.55e-08 ***
OVSlt-OVSel               -2.12304    0.37382  -5.679 1.35e-08 ***
lang_exp                   0.03371    0.01555   2.167   0.0302 *
L_edu                      0.19698    0.43534   0.452   0.6509
no_older_siblings         -0.38976    0.40497  -0.962   0.3358
OVSlt-OVSel:lang_exp       0.01566    0.01984   0.789   0.4299
OVSlt-OVSel:L_edu          0.93698    0.58993   1.588   0.1122
OVSlt-OVSel:no_older_sib   0.92775    0.49705   1.867   0.0620 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.28 Output statistical model Table 24

Repeated contrasts used for condition. Only the monolingual group is included. Random intercepts: participants and items. Random slope: individual adjustment of condition for each participant. Continuous predictors: number of older siblings and parents' education. Correct responses as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
 Family: binomial  ( logit )
Formula: acc~1+cond*(L_edu+no_older_sib)+(1+cond|id)+(1|item)
 Data: df1
Control: glmerControl(calc.derivs = FALSE)

     AIC      BIC   logLik deviance df.resid
   379.7    417.4   -179.8    359.7      310

Scaled residuals:
     Min      1Q  Median     3Q      Max
 -2.4613 -0.6138  0.2196  0.5904  2.6102

Random effects:
 Groups   Name        Variance  Std.Dev.  Corr
 item   (Intercept)    0.2033   0.4509
 id     (Intercept)    1.4877   1.2197
        OVSlt-OVSel    0.9665   0.9831   0.50
Number of obs: 320, groups:  item, 32; id, 20

Fixed effects:
                           Estimate Std. Error z value Pr(>|z|)
(Intercept)                0.101403   0.318073   0.319   0.7499
OVSlt-OVSel               -1.740218   0.393596  -4.421 9.81e-06***
L_edu                      0.896070   0.505218   1.774   0.0761.
no_older_siblings         -0.497828   0.524444  -0.949   0.3425
OVSlt-OVSel:L_edu          0.082856   0.585869   0.141   0.8875
OVSlt-OVSel:no_older_sib   0.007082   0.605908   0.012   0.9907
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.29 Output statistical model Table 26

Repeated contrasts used for condition. Language group coded with treatment contrast: bilinguals in intercept. Random intercepts: participants and items. Random slope: individual adjustment of condition for each participant. Continuous predictors: language knowledge, memory, and reasoning ability. Correct responses as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Family: binomial  ( logit )
Formula:
acc~1+lang_group*cond*(lang_knowl+mem+reas)+(1+cond|id)+(1|item)
Data: d
Control: glmerControl(calc.derivs = FALSE)

     AIC      BIC   logLik deviance df.resid
  1067.2   1165.2   -513.6   1027.2      972

Scaled residuals:
     Min      1Q    Median     3Q      Max
 -2.7676 -0.5658   -0.2606  0.5662  7.7819

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 id      (Intercept) 1.1448    1.0699
         OVSlt-OVSel 0.2001    0.4473   0.16
 item    (Intercept) 0.1264    0.3555
Number of obs: 992, groups:  id, 62; item, 32

Fixed effects:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)              -1.437099   0.254501  -5.647 1.64e-08***
mo                        1.370551   0.331355   4.136 3.53e-05***
OVSlt-OVSel              -2.294562   0.340080  -6.747 1.51e-11***
lang_knowl                0.052572   0.028968   1.815  0.06955.
mem                       0.077378   0.035126   2.203  0.02760*
reas                     -0.054125   0.027902  -1.940  0.05240.
mo:OVSlt-OVSel            0.678901   0.389443   1.743  0.08129.
mo:lang_knowl            -0.010735   0.045609  -0.235  0.81392
mo:mem                   -0.096915   0.049801  -1.946  0.05165.
mo:reas                   0.076209   0.038940   1.957  0.05034.
                           Estimate Std. Error z value Pr(>|z|)
OVSlt-OVSel:lang_knowl    -0.098493   0.034442  -2.860  0.00424**
OVSlt-OVSel:mem            0.093203   0.043664   2.135  0.03280*
OVSlt-OVSel:reas          -0.015713   0.034528  -0.455  0.64904
mo:OVSlt-OVSel:lang_knowl  0.051079   0.050202   1.017  0.30893
mo:OVSlt-OVSel:mem        -0.064677   0.056996  -1.135  0.25648
mo:OVSlt-OVSel:reas       -0.009582   0.044546  -0.215  0.82969
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C.30 Output statistical model Table 27

Repeated contrasts used for condition. Language group coded with treatment contrast: monolinguals in the intercept. Random intercepts: participants and items. Random slope: individual adjustment of condition effect for each participant. Continuous predictors: language knowledge, memory, and reasoning ability. Correct responses as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Family: binomial  ( logit )
Formula:
acc~1+lang_group*cond*(lang_knowl+mem+reas)+(1+cond|id)+(1|item)
Data: d
Control: glmerControl(calc.derivs = FALSE)

     AIC      BIC   logLik deviance df.resid
  1067.2   1165.2   -513.6   1027.2      972

Scaled residuals:
    Min      1Q   Median      3Q     Max
-2.7560 -0.5658  -0.2603  0.5655  7.7608

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 id       (Intercept) 1.1493    1.0721
          OVSlt-OVSel 0.2125    0.4610   0.16
 item     (Intercept) 0.1272    0.3567
Number of obs: 992, groups:  id, 62; item, 32

Fixed effects:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)              -0.067202   0.230543  -0.291   0.7707
bi                       -1.372811   0.331934  -4.136 3.54e-05***
OVSlt-OVSel              -1.615942   0.261167  -6.187 6.12e-10***
lang_knowl                0.041887   0.035232   1.189   0.2345
mem                      -0.019389   0.035377  -0.548   0.5836
reas                      0.022297   0.027154   0.821   0.4116
bi:OVSlt-OVSel           -0.682149   0.390892  -1.745   0.0810.
bi:lang_knowl             0.010652   0.045685   0.233   0.8156
bi:mem                    0.096847   0.049886   1.941   0.0522.
bi:reas                  -0.076330   0.039004  -1.957   0.0503.
                           Estimate Std. Error z value Pr(>|z|)
OVSlt-OVSel:lang_knowl    -0.047472   0.036477  -1.301   0.1931
OVSlt-OVSel:mem            0.028577   0.036855   0.775   0.4381
OVSlt-OVSel:reas          -0.025033   0.028068  -0.892   0.3725
bi:OVSlt-OVSel:lang_knowl -0.051031   0.050393  -1.013   0.3112
bi:OVSlt-OVSel:mem         0.064968   0.057212   1.136   0.2561
bi:OVSlt-OVSel:reas        0.009422   0.044700   0.211   0.8331
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

D PASSIVE: material

The four versions of the passive experiment.

Passive Version 1

Date: Code:

Date of birth: Sex:

pos	item	sentence	targ	com
1	20	Christina_Intro	-	
2	201	Schau mal, ein Ball. Welche Farbe hat der Rahmen?	gelb	
3	203	Schau mal, eine Banane. Welche Farbe hat der Rahmen?	blau	
4	205	Schau mal, die Ameise weint. Welche Farbe hat der Rahmen?	blau	
5	207	Die Schnecke rutscht. Welche Farbe hat der Rahmen?	gelb	
6	209	Die Biene schneidet Papier	gelb	
7	211	Die Kuh fährt das Auto.	blau	
8	213	Christina_los geht's	-	
9	22231a	Kürzlich hat der Dachs das Schaf getreten	gelb	
10	268121	Letztens wurde das Pferd von dem Tiger gerufen	blau	
11	23451b	Der Frosch hat täglich das Eichhörnchen gekitzelt	blau	
12	283161	Das Reh wurde einmal von dem Flamingo gewaschen	gelb	
13	24571b	Der Hai hat damals das Zebra geschlagen	gelb	
14	257101	Vorhin wurde das Schwein von dem Pfau geschoben	blau	
15	272141	Das Känguru wurde kürzlich von dem Esel getreten	gelb	
16	21111a	Gestern hat der Panther das Nilpferd gefangen	blau	
17	214	Christina_weiter so	-	
18	267111	Das Pferd hat vorhin den Tiger geschoben	gelb	
19	22541d	Der Dachs wurde damals von dem Schaf geschlagen	gelb	
20	24471a	Täglich hat der Hai das Zebra gekitzelt	blau	
21	23351a	Einmal hat der Frosch das Eichhörnchen gewaschen	blau	
22	271141	Gestern wurde das Känguru von dem Esel gefangen	blau	
23	25691b	Das Schwein hat neulich den Pfau gemessen	gelb	
24	21421d	Der Panther wurde täglich von dem Nilpferd gekitzelt	blau	
25	282161	Kürzlich wurde das Reh von dem Flamingo getreten	blau	
26	215	Christina_die Hälfte	-	
27	25591a	Damals hat das Schwein den Pfau geschlagen	gelb	
28	23661d	Der Frosch wurde neulich von dem Eichhörnchen	gelb	
29	281151	Das Reh hat gestern den Flamingo gefangen	blau	
30	21321c	Einmal wurde der Panther von dem Nilpferd gewaschen	blau	
31	266111	Neulich hat das Pferd den Tiger gemessen	gelb	
32	22441c	Täglich wurde der Dachs von dem Schaf gekitzelt	gelb	
33	278131	Das Känguru hat letztens den Esel gerufen	blau	
34	24781d	Der Hai wurde vorhin von dem Zebra geschoben	blau	
35	216	Christina_bald fertig	-	
36	23561c	Damals wurde der Frosch von dem Eichhörnchen	gelb	
37	258101	Das Schwein wurde letztens von dem Pfau gerufen	gelb	
38	21211b	Der Panther hat kürzlich das Nilpferd getreten	gelb	
39	24681c	Neulich wurde der Hai von dem Zebra gemessen	blau	
40	261121	Das Pferd wurde gestern von dem Tiger gefangen	gelb	
41	288151	Letztens hat das Reh den Flamingo gerufen	gelb	
42	277131	Vorhin hat das Känguru den Esel geschoben	blau	
43	22331b	Der Dachs hat einmal das Schaf gewaschen	blau	
44	217	Christina_Applaus	-	

Passive Version 2

Date: Code:

Date of birth: Sex:

pos	item	sentence	targ	com
1	20	Christina_Intro	-	
2	201	Schau mal, ein Ball. Welche Farbe hat der Rahmen?	gelb	
3	203	Schau mal, eine Banane. Welche Farbe hat der Rahmen?	blau	
4	205	Schau mal, die Ameise weint. Welche Farbe hat der Rahmen?	blau	
5	207	Die Schnecke rutscht. Welche Farbe hat der Rahmen?	gelb	
6	209	Die Biene schneidet Papier	gelb	
7	211	Die Kuh fährt das Auto.	blau	
8	213	Christina_los geht's	-	
9	24472b	Der Hai hat täglich das Zebra gekitzelt	gelb	
10	258102	Letztens wurde das Schwein von dem Pfau gerufen	blau	
11	21212a	Kürzlich hat der Panther das Nilpferd getreten	gelb	
12	261122	Gestern wurde das Pferd von dem Tiger gefangen	gelb	
13	282162	Das Reh wurde kürzlich von dem Flamingo getreten	blau	
14	22332a	Einmal hat der Dachs das Schaf gewaschen	blau	
15	271142	Das Känguru wurde gestern von dem Esel gefangen	blau	
16	23352b	Der Frosch hat einmal das Eichhörnchen gewaschen	gelb	
17	214	Christina_weiter so	-	
18	24572a	Damals hat der Hai das Zebra geschlagen	gelb	
19	272142	Kürzlich wurde das Känguru von dem Esel getreten	blau	
20	22442d	Der Dachs wurde täglich von dem Schaf gekitzelt	gelb	
21	283162	Einmal wurde das Reh von dem Flamingo gewaschen	gelb	
22	25592b	Das Schwein hat damals den Pfau geschlagen	blau	
23	21322d	Der Panther wurde einmal von dem Nilpferd gewaschen	blau	
24	266112	Das Pferd hat neulich den Tiger gemessen	gelb	
25	23452a	Täglich hat der Frosch das Eichhörnchen gekitzelt	gelb	
26	215	Christina_die Hälfte	-	
27	288152	Das Reh hat letztens den Flamingo gerufen	gelb	
28	23562d	Der Frosch wurde damals von dem Eichhörnchen	blau	
29	25692a	Neulich hat das Schwein den Pfau gemessen	blau	
30	277132	Das Känguru hat vorhin den Esel geschoben	blau	
31	22542c	Damals wurde der Dachs von dem Schaf geschlagen	gelb	
32	267112	Vorhin hat das Pferd den Tiger geschoben	blau	
33	21422c	Täglich wurde der Panther von dem Nilpferd gekitzelt	blau	
34	24682d	Der Hai wurde neulich von dem Zebra gemessen	gelb	
35	216	Christina_bald fertig	-	
36	278132	Letztens hat das Känguru den Esel gerufen	blau	
37	24782c	Vorhin wurde der Hai von dem Zebra geschoben	blau	
38	268122	Das Pferd wurde letztens von dem Tiger gerufen	blau	
39	21112b	Der Panther hat gestern das Nilpferd gefangen	gelb	
40	257102	Das Schwein wurde vorhin von dem Pfau geschoben	gelb	
41	281152	Gestern hat das Reh den Flamingo gefangen	gelb	
42	23662c	Neulich wurde der Frosch von dem Eichhörnchen	blau	
43	22232b	Der Dachs hat kürzlich das Schaf getreten	gelb	
44	217	Christina_Applaus	-	

Passive Version 3

Date: Code:

Date of birth: Sex:

pos	item	sentence	targ	com
1	20	Christina_Intro	-	
2	201	Schau mal, ein Ball. Welche Farbe hat der Rahmen?	gelb	
3	203	Schau mal, eine Banane. Welche Farbe hat der Rahmen?	blau	
4	205	Schau mal, die Ameise weint. Welche Farbe hat der Rahmen?	blau	
5	207	Die Schnecke rutscht. Welche Farbe hat der Rahmen?	gelb	
6	209	Die Biene schneidet Papier	gelb	
7	211	Die Kuh fährt das Auto.	blau	
8	213	Christina_los geht's	-	
9	23653b	Der Frosch hat neulich das Eichhörnchen gemessen	gelb	
10	278143	Das Känguru wurde letztens von dem Esel gerufen	blau	
11	266123	Neulich wurde das Pferd von dem Tiger gemessen	blau	
12	22433a	Täglich hat der Dachs das Schaf gekitzelt	gelb	
13	255103	Damals wurde das Schwein von dem Pfau geschlagen	gelb	
14	21313a	Einmal hat der Panther das Nilpferd gewaschen	blau	
15	281163	Das Reh wurde gestern von dem Flamingo gefangen	gelb	
16	24773b	Der Hai hat vorhin das Zebra geschoben	gelb	
17	214	Christina_weiter so	-	
18	21223d	Der Panther wurde kürzlich von dem Nilpferd getreten	blau	
19	25893b	Das Schwein hat letztens den Pfau gerufen	gelb	
20	23553a	Damals hat der Frosch das Eichhörnchen geschlagen	gelb	
21	261113	Das Pferd hat gestern den Tiger gefangen	gelb	
22	277143	Vorhin wurde das Känguru von dem Esel geschoben	blau	
23	22343d	Der Dachs wurde einmal von dem Schaf gewaschen	gelb	
24	288163	Letztens wurde das Reh von dem Flamingo gerufen	blau	
25	24673a	Neulich hat der Hai das Zebra gemessen	blau	
26	215	Christina_die Hälfte	-	
27	23463d	Der Frosch wurde täglich von dem Eichhörnchen gekitzelt	blau	
28	21123c	Gestern wurde der Panther von dem Nilpferd gefangen	gelb	
29	272133	Das Känguru hat kürzlich den Esel getreten	blau	
30	25793a	Vorhin hat das Schwein den Pfau geschoben	blau	
31	283153	Das Reh hat einmal den Flamingo gewaschen	gelb	
32	22243c	Kürzlich wurde der Dachs von dem Schaf getreten	gelb	
33	268113	Letztens hat das Pferd den Tiger gerufen	blau	
34	24583d	Der Hai wurde damals von dem Zebra geschlagen	gelb	
35	216	Christina_bald fertig	-	
36	21413b	Der Panther hat täglich das Nilpferd gekitzelt	blau	
37	256103	Das Schwein wurde neulich von dem Pfau gemessen	gelb	
38	282153	Kürzlich hat das Reh den Flamingo getreten	blau	
39	23363c	Einmal wurde der Frosch von dem Eichhörnchen	blau	
40	271133	Gestern hat das Känguru den Esel gefangen	gelb	
41	22533b	Der Dachs hat damals das Schaf geschlagen	blau	
42	267123	Das Pferd wurde vorhin von dem Tiger geschoben	gelb	
43	24483c	Täglich wurde der Hai von dem Zebra gekitzelt	blau	
44	217	Christina_Applaus	-	

Passive Version 4

Date: Code:

Date of birth: Sex:

pos	item	sentence	targ	com
1	20	Christina_Intro	-	
2	201	Schau mal, ein Ball. Welche Farbe hat der Rahmen?	gelb	
3	203	Schau mal, eine Banane. Welche Farbe hat der Rahmen?	blau	
4	205	Schau mal, die Ameise weint. Welche Farbe hat der Rahmen?	blau	
5	207	Die Schnecke rutscht. Welche Farbe hat der Rahmen?	gelb	
6	209	Die Biene schneidet Papier	gelb	
7	211	Die Kuh fährt das Auto.	blau	
8	213	Christina_los geht's	-	
9	23554b	Der Frosch hat damals das Eichhörnchen geschlagen	gelb	
10	277144	Das Känguru wurde vorhin von dem Esel geschoben	blau	
11	22534a	Damals hat der Dachs das Schaf geschlagen	blau	
12	24674b	Der Hai hat neulich das Zebra gemessen	gelb	
13	267124	Vorhin wurde das Pferd von dem Tiger geschoben	blau	
14	21414a	Täglich hat der Panther das Nilpferd gekitzelt	gelb	
15	288164	Das Reh wurde letztens von dem Flamingo gerufen	gelb	
16	256104	Neulich wurde das Schwein von dem Pfau gemessen	blau	
17	214	Christina_weiter so	-	
18	23654a	Neulich hat der Frosch das Eichhörnchen gemessen	gelb	
19	21124d	Der Panther wurde gestern von dem Nilpferd gefangen	blau	
20	278144	Letztens wurde das Känguru von dem Esel gerufen	gelb	
21	22244d	Der Dachs wurde kürzlich von dem Schaf getreten	blau	
22	268114	Das Pferd hat letztens den Tiger gerufen	blau	
23	24774a	Vorhin hat der Hai das Zebra geschoben	gelb	
24	281164	Gestern wurde das Reh von dem Flamingo gefangen	gelb	
25	25794b	Das Schwein hat vorhin den Pfau geschoben	blau	
26	215	Christina_die Hälfte	-	
27	22344c	Einmal wurde der Dachs von dem Schaf gewaschen	gelb	
28	261114	Gestern hat das Pferd den Tiger gefangen	blau	
29	282154	Das Reh hat kürzlich den Flamingo getreten	blau	
30	24484d	Der Hai wurde täglich von dem Zebra gekitzelt	gelb	
31	25894a	Letztens hat das Schwein den Pfau gerufen	blau	
32	271134	Das Känguru hat gestern den Esel gefangen	gelb	
33	23364d	Der Frosch wurde einmal von dem Eichhörnchen	gelb	
34	21224c	Kürzlich wurde der Panther von dem Nilpferd getreten	blau	
35	216	Christina_bald fertig	-	
36	22434b	Der Dachs hat täglich das Schaf gekitzelt	gelb	
37	266124	Das Pferd wurde neulich von dem Tiger gemessen	blau	
38	272134	Kürzlich hat das Känguru den Esel getreten	blau	
39	255104	Das Schwein wurde damals von dem Pfau geschlagen	gelb	
40	23464c	Täglich wurde der Frosch von dem Eichhörnchen gekitzelt	blau	
41	21314b	Der Panther hat einmal das Nilpferd gewaschen	blau	
42	283154	Einmal hat das Reh den Flamingo gewaschen	gelb	
43	24584c	Damals wurde der Hai von dem Zebra geschlagen	gelb	
44	217	Christina_Applaus	-	

E PASSIVE: statistics

E.1 Output statistical model Table 32

Treatment contrast for language group and condition. Monolinguals and PAS PF in intercept. Random intercepts: participants and items. Random slope: individual adjustment of condition effect for each participant. Correct resp as dependent variable.

```
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) ['glmerMod']
Family: binomial  ( logit )
Formula: acc ~ 1 + lang_group * cond + (1 + cond|id) + (1|item)
Data: d
Control: glmerControl(calc.derivs = FALSE)

     AIC      BIC   logLik deviance df.resid
  1397.1   1500.0   -679.5   1359.1     1645

Scaled residuals:
    Min      1Q  Median      3Q     Max
-5.3743  0.2498  0.3582  0.4384  1.0403

Random effects:
Groups Name        Variance Std.Dev.  Corr
 item  (Intercept)  0.04525  0.2127
 id    (Intercept)  0.39993  0.6324
        PASMF       0.31896  0.5648  -0.60
        ACTPF       0.72767  0.8530  -0.02  0.81
        ACTMF       0.29595  0.5440  -0.85  0.93  0.55
Number of obs: 1664, groups:  item, 64; id, 52

Fixed effects:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)       2.36443    0.27308   8.658   <2e-16***
bili             -0.66615    0.35292  -1.888   0.0591.
PASMF            -0.58383    0.32750  -1.783   0.0746.
ACTPF             0.25713    0.39381   0.653   0.5138
ACTMF            -0.60299    0.33048  -1.825   0.0681.
bili:PASMF       -0.06553    0.42523  -0.154   0.8775
bili:ACTPF        0.64619    0.53115   1.217   0.2238
bili:ACTMF        0.66063    0.43721   1.511   0.1308
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.2 Output statistical model Table 33

Treatment contrast for language group and condition. Bilinguals and PAS MF in intercept. Random intercepts: participants and items. Random slope: individual adjustment of condition effect for each participant. Correct resp as dependent variable.

```
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) ['glmerMod']
Family: binomial  ( logit )
Formula: acc ~ 1 + lang_group * cond + (1 + cond|id) + (1|item)
Data: d
Control: glmerControl(calc.derivs = FALSE)

     AIC      BIC   logLik deviance df.resid
  1397.0   1500.0   -679.5   1359.0     1645

Scaled residuals:
    Min      1Q  Median      3Q     Max
-5.3470  0.2474  0.3588  0.4373  1.0390

Random effects:
Groups Name        Variance Std.Dev.  Corr
 item  (Intercept)  0.04519  0.2126
 id    (Intercept)  0.28546  0.5343
        PASPF       0.34064  0.5836   -0.33
        ACTPF       0.26266  0.5125    0.99 -0.21
        ACTMF       0.04125  0.2031   -1.00  0.29 -1.00
Number of obs: 1664, groups:  item, 64; id, 52

Fixed effects:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)       1.04819    0.19360   5.414 6.16e-08***
mono              0.72923    0.29075   2.508  0.01214*
PASPF             0.65774    0.27243   2.414  0.01576*
ACTPF             1.54822    0.32092   4.824 1.40e-06***
ACTMF             0.70614    0.25831   2.734  0.00626**
mono:PASPF       -0.05779    0.42792  -0.135  0.89257
mono:ACTPF       -0.71406    0.46124  -1.548  0.12159
mono:ACTMF       -0.72548    0.37425  -1.938  0.05256.
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.3 Output statistical model Table 34

Treatment contrast for language group and condition. Monolinguals and PAS PF in intercept. Random intercepts: participants and items. Model output for first time window PF. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "PF")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 1902.5

Scaled residuals:
     Min       1Q   Median       3Q      Max
-1.18895 -1.03962  0.03365  1.03324  1.18925

Random effects:
 Groups   Name        Variance  Std.Dev.
 item     (Intercept) 1.126e-05 0.003355
 id       (Intercept) 0.000e+00 0.000000
 Residual             2.081e-01 0.456229
Number of obs: 1475, groups:  item, 64; id, 52

Fixed effects:
               Estimate Std. Error         df t value Pr(>|t|)
(Intercept)     0.45762    0.03285  647.78669  13.933   <2e-16***
bili            0.01670    0.04681 1443.88144   0.357    0.721
PASMF           0.00239    0.04727 1423.48091   0.051    0.960
ACTPF           0.07101    0.04708  669.27200   1.508    0.132
ACTMF           0.04635    0.04735  678.94863   0.979    0.328
bili:PASMF      0.06026    0.06697 1460.64471   0.900    0.368
bili:ACTPF     -0.01857    0.06674 1440.25211  -0.278    0.781
bili:ACTMF      0.02145    0.06693 1444.35767   0.320    0.749
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.4 Output statistical model Table 34

Treatment contrast for language group and condition. Monolinguals and PAS PF in intercept. Random intercepts: participants and items. Model output for time window AUX. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "AUX")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 1995

Scaled residuals:
    Min      1Q  Median     3Q    Max
-1.2052 -1.0050  0.3197 0.9756 1.1414

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept) 0.0000   0.0000
 id       (Intercept) 0.0000   0.0000
 Residual             0.2331   0.4828
Number of obs: 1422, groups:  item, 64; id, 52

Fixed effects:
               Estimate Std. Error         df t value Pr(>|t|)
(Intercept)     0.44895    0.03531 1414.00000  12.716   <2e-16***
bili            0.03625    0.05020 1414.00000   0.722   0.4704
PASMF           0.03412    0.05085 1414.00000   0.671   0.5023
ACTPF           0.08746    0.05027 1414.00000   1.740   0.0821.
ACTMF           0.03063    0.05116 1414.00000   0.599   0.5495
bili:PASMF      0.06255    0.07180 1414.00000   0.871   0.3838
bili:ACTPF     -0.04366    0.07175 1414.00000  -0.609   0.5429
bili:ACTMF      0.05274    0.07238 1414.00000   0.729   0.4663
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.5 Output statistical model Table 34

Treatment contrast for language group and condition. Monolinguals and PAS PF in intercept. Random intercepts: participants and items. Model output for time window MF1. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "MF1")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 2011.2

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.1964 -1.0304  0.1419  0.9986  1.2323

Random effects:
 Groups   Name        Variance  Std.Dev.
 item     (Intercept) 0.0000    0.0000
 id       (Intercept) 0.0000    0.0000
 Residual             0.2207    0.4698
Number of obs: 1492, groups:  item, 64; id, 52

Fixed effects:
              Estimate Std. Error         df t value Pr(>|t|)
(Intercept)    0.42104    0.03445 1484.00000  12.222   <2e-16***
bili           0.05760    0.04859 1484.00000   1.185   0.2360
PASMF          0.08281    0.04878 1484.00000   1.698   0.0898.
ACTPF          0.10980    0.04852 1484.00000   2.263   0.0238*
ACTMF          0.06307    0.04865 1484.00000   1.296   0.1951
bili:PASMF    -0.01797    0.06872 1484.00000  -0.262   0.7937
bili_ACTPF    -0.03261    0.06876 1484.00000  -0.474   0.6354
bili:ACTMF     0.02039    0.06885 1484.00000   0.296   0.7671
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.6 Output statistical model Table 34

Treatment contrast for language group and condition. Monolinguals and PAS PF in intercept. Random intercepts: participants and items. Model output for time window MF2. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "MF2")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 1827.8

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.4818 -1.2571  0.4456  0.9254  1.0498

Random effects:
 Groups   Name        Variance  Std.Dev.
 item     (Intercept) 0.0002644 0.01626
 id       (Intercept) 0.0000000 0.00000
 Residual             0.1890143 0.43476
Number of obs: 1530, groups:  item, 64; id, 52

Fixed effects:
              Estimate Std. Error        df t value Pr(>|t|)
(Intercept)  5.872e-01  3.143e-02 6.731e+02  18.682   <2e-16***
bili        -1.417e-03  4.410e-02 1.498e+03  -0.032    0.974
PASMF        5.874e-03  4.421e-02 1.478e+03   0.133    0.894
ACTPF        5.134e-02  4.433e-02 6.682e+02   1.158    0.247
ACTMF       -3.722e-02  4.451e-02 6.784e+02  -0.836    0.403
bili:PASMF  -4.344e-02  6.258e-02 1.516e+03  -0.694    0.488
bili:ACTPF  -2.433e-03  6.266e-02 1.496e+03  -0.039    0.969
bili:ACTMF   5.297e-02  6.295e-02 1.498e+03   0.841    0.400
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.7 Output statistical model Table 34

Treatment contrast for language group and condition. Monolinguals and PAS PF in intercept. Random intercepts: participants and items. Model output for time window V. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "V")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 1818.2

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.8484 -1.2651  0.5991  0.7418  1.1202

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept) 0.001714 0.04140
 id       (Intercept) 0.005293 0.07275
 Residual             0.185151 0.43029
Number of obs: 1514, groups:  item, 64; id, 52

Fixed effects:
                Estimate Std. Error        df t value Pr(>|t|)
(Intercept)    6.867e-01  3.492e-02  2.733e+02  19.666  <2e-16***
bili          -3.533e-02  4.847e-02  3.097e+02  -0.729   0.467
PASMF         -3.361e-03  4.392e-02  1.421e+03  -0.077   0.939
ACTPF          2.509e-02  4.505e-02  5.488e+02   0.557   0.578
ACTMF         -3.209e-02  4.517e-02  5.532e+02  -0.711   0.478
bili:PASMF    -3.019e-02  6.243e-02  1.451e+03  -0.484   0.629
bili:ACTPF     7.872e-03  6.247e-02  1.433e+03   0.126   0.900
bili:ACTMF     7.070e-02  6.289e-02  1.437e+03   1.124   0.261
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.8 Output statistical model Table 34

Treatment contrast for language group and condition. Monolinguals and PAS PF in intercept. Random intercepts: participants and items. Model output for time window Sil1. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "Sil1")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 1685.3

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.0391 -1.0379  0.5436  0.7065  1.2181

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept) 0.000000 0.00000
 id       (Intercept) 0.007493 0.08656
 Residual             0.167802 0.40964
Number of obs: 1529, groups:  item, 64; id, 52

Fixed effects:
              Estimate Std. Error        df t value Pr(>|t|)
(Intercept)  7.102e-01  3.412e-02  2.500e+02  20.819   <2e-16***
bili        -1.859e-02  4.805e-02  2.466e+02  -0.387    0.699
PASMF        3.349e-04  4.185e-02  1.476e+03   0.008    0.994
ACTPF        3.331e-02  4.184e-02  1.475e+03   0.796    0.426
ACTMF        2.146e-03  4.194e-02  1.474e+03   0.051    0.959
bili:PASMF  -5.119e-02  5.890e-02  1.475e+03  -0.869    0.385
bili:ACTPF  -4.170e-02  5.927e-02  1.476e+03  -0.704    0.482
bili:ACTMF   4.336e-02  5.933e-02  1.475e+03   0.731    0.465
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.9 Output statistical model Table 34

Treatment contrast for language group and condition. Monolinguals and PAS PF in intercept. Random intercepts: participants and items. Model output for time window Sil2. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "Sil2")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 1622.7

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.9779 -0.9878  0.5891  0.7443  1.1446

Random effects:
 Groups   Name        Variance  Std.Dev.
 item     (Intercept) 4.443e-05 0.006665
 id       (Intercept) 4.171e-03 0.064582
 Residual             1.670e-01 0.408702
Number of obs: 1490, groups:  item, 64; id, 52

Fixed effects:
              Estimate Std. Error        df t value Pr(>|t|)
(Intercept)  7.083e-01  3.282e-02  2.996e+02  21.582   <2e-16***
bili        -1.148e-02  4.590e-02  3.426e+02  -0.250    0.803
PASMF       -4.930e-02  4.267e-02  1.403e+03  -1.155    0.248
PASPF        2.546e-02  4.233e-02  6.379e+02   0.601    0.548
ACTMF       -7.294e-03  4.258e-02  6.491e+02  -0.171    0.864
Bili:PASMF  -4.609e-02  6.005e-02  1.431e+03  -0.768    0.443
Bili:ACTPF  -1.056e-02  5.979e-02  1.408e+03  -0.177    0.860
Bili:ACTMF   2.003e-02  5.983e-02  1.412e+03   0.335    0.738
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.10 Output statistical model Table 34

Treatment contrast for language group and condition. Monolinguals and PAS PF in intercept. Random intercepts: participants and items. Model output for time window Sil3. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "Sil3")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 1571.3

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.8020 -1.2463  0.6315  0.8036  1.0690

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept) 0.000000 0.00000
 id       (Intercept) 0.002065 0.04544
 Residual             0.179228 0.42335
Number of obs: 1364, groups:  item, 64; id, 52

Fixed effects:
             Estimate Std. Error        df t value Pr(>|t|)
(Intercept)   0.62807    0.03380  443.02999  18.581   <2e-16***
bili          0.02535    0.04742  432.10011   0.535   0.5932
PASMF        -0.03990    0.04659 1319.66440  -0.857   0.3919
ACTPF         0.09385    0.04574 1312.63430   2.052   0.0404*
ACTMF         0.04761    0.04562 1317.52132   1.044   0.2969
bili:PASMF   -0.01186    0.06521 1320.62604  -0.182   0.8557
bili:ACTPF   -0.04966    0.06500 1314.79028  -0.764   0.4450
bili:ACTMF   -0.05245    0.06424 1316.84045  -0.816   0.4144
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.11 Output statistical model Table 35

Treatment contrast for language group and condition. Bilinguals and PAS MF in intercept. Random intercepts: participants and items. Model output for first time window PF. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "PF")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 1902.5

Scaled residuals:
     Min      1Q   Median       3Q      Max
-1.18895 -1.03962  0.03365  1.03324  1.18925

Random effects:
 Groups   Name         Variance  Std.Dev.
 item     (Intercept) 1.126e-05 0.003355
 id       (Intercept) 0.000e+00 0.000000
 Residual             2.081e-01 0.456229
Number of obs: 1475, groups:  item, 64; id, 52

Fixed effects:
               Estimate Std. Error         df t value Pr(>|t|)
(Intercept)   5.370e-01  3.373e-02  6.496e+02  15.919   <2e-16***
mono         -7.697e-02  4.789e-02  1.443e+03  -1.607    0.108
PASPF        -6.265e-02  4.744e-02  1.465e+03  -1.321    0.187
ACTPF        -1.022e-02  4.757e-02  6.445e+02  -0.215    0.830
ACTMF         5.147e-03  4.757e-02  6.423e+02   0.108    0.914
mono:PASPF    6.026e-02  6.697e-02  1.461e+03   0.900    0.368
mono:ACTPF    7.884e-02  6.750e-02  1.440e+03   1.168    0.243
mono:ACTMF    3.882e-02  6.769e-02  1.444e+03   0.573    0.566
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.12 Output statistical model Table 35

Treatment contrast for language group and condition. Bilinguals and PAS MF in intercept. Random intercepts: participants and items. Model output for time window AUX. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "AUX")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 1995

Scaled residuals:
    Min      1Q   Median      3Q     Max
-1.2052  -1.0050   0.3197  0.9756  1.1414

Random effects:
 Groups   Name      Variance  Std.Dev.
 item     (Intercept) 0.0000   0.0000
 id       (Intercept) 0.0000   0.0000
 Residual             0.2331   0.4828
Number of obs: 1422, groups:  item, 64; id, 52

Fixed effects:
               Estimate Std. Error         df t value Pr(>|t|)
(Intercept)   5.819e-01  3.599e-02  1.414e+03  16.169   <2e-16***
mono         -9.880e-02  5.133e-02  1.414e+03  -1.925   0.0545.
PASPF        -9.667e-02  5.068e-02  1.414e+03  -1.907   0.0567.
ACTPF        -5.287e-02  5.140e-02  1.414e+03  -1.029   0.3039
ACTMF        -1.330e-02  5.140e-02  1.414e+03  -0.259   0.7959
mono:PASPF    6.255e-02  7.180e-02  1.414e+03   0.871   0.3838
mono:ACTPF    1.062e-01  7.255e-02  1.414e+03   1.464   0.1434
mono:ACTMF    9.805e-03  7.317e-02  1.414e+03   0.134   0.8934
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.13 Output statistical model Table 35

Treatment contrast for language group and condition. Bilinguals and PAS MF in intercept. Random intercepts: participants and items. Model output for time window MF1. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "MF1")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 2011.2

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.1964 -1.0304  0.1419  0.9986  1.2323

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept) 0.0000   0.0000
 id       (Intercept) 0.0000   0.0000
 Residual             0.2207   0.4698
Number of obs: 1492, groups:  item, 64; id, 52

Fixed effects:
             Estimate Std. Error         df t value Pr(>|t|)
(Intercept)   0.54348    0.03417 1484.00000  15.903   <2e-16 ***
mono         -0.03962    0.04859 1484.00000  -0.815    0.415
PASPF        -0.06484    0.04839 1484.00000  -1.340    0.181
ACTPF         0.01235    0.04866 1484.00000   0.254    0.800
ACTMF         0.01862    0.04866 1484.00000   0.383    0.702
mono:PASPF   -0.01797    0.06872 1484.00000  -0.262    0.794
mono:ACTPF    0.01463    0.06877 1484.00000   0.213    0.832
mono:ACTMF   -0.03837    0.06886 1484.00000  -0.557    0.577
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.14 Output statistical model Table 35

Treatment contrast for language group and condition. Bilinguals and PAS MF in intercept. Random intercepts: participants and items. Model output for time window MF2. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "MF2")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 1827.8

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.4818 -1.2571  0.4456  0.9254  1.0498

Random effects:
 Groups   Name        Variance  Std.Dev.
 item     (Intercept) 0.0002644 0.01626
 id       (Intercept) 0.0000000 0.00000
 Residual             0.1890143 0.43476
Number of obs: 1530, groups:  item, 64; id, 52

Fixed effects:
             Estimate Std. Error         df t value Pr(>|t|)
(Intercept)   0.54817    0.03168  648.07203  17.303   <2e-16 ***
mono          0.04486    0.04439 1498.26449   1.011    0.312
PASPF         0.03757    0.04429 1518.29282   0.848    0.396
ACTPF         0.08648    0.04498  656.02463   1.922    0.055.
ACTMF         0.05332    0.04523  662.04533   1.179    0.239
mono:PASPF   -0.04344    0.06258 1515.54909  -0.694    0.488
mono:ACTPF   -0.04101    0.06286 1496.26051  -0.652    0.514
mono:ACTMF   -0.09641    0.06315 1497.48783  -1.527    0.127
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.15 Output statistical model Table 35

Treatment contrast for language group and condition. Bilinguals and PAS MF in intercept. Random intercepts: participants and items. Model output for the time window V. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "V")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 1818.2

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.8484 -1.2651  0.5991  0.7418  1.1202

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept) 0.001714 0.04140
 id       (Intercept) 0.005293 0.07275
 Residual             0.185151 0.43029
Number of obs: 1514, groups:  item, 64; id, 52

Fixed effects:
              Estimate Std. Error         df t value Pr(>|t|)
(Intercept)    0.61779    0.03517  276.32621  17.567   <2e-16***
mono           0.06552    0.04853  309.68526   1.350    0.178
PASPF          0.03355    0.04430 1454.99024   0.757    0.449
ACTPF          0.06650    0.04568  545.79820   1.456    0.146
ACTMF          0.07215    0.04610  555.27688   1.565    0.118
mono:PASPF    -0.03019    0.06243 1451.20352  -0.484    0.629
mono:ACTPF    -0.03806    0.06252 1432.13568  -0.609    0.543
mono:ACTMF    -0.10088    0.06291 1434.78433  -1.604    0.109
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.16 Output statistical model Table 35

Treatment contrast for language group and condition. Bilinguals and PAS MF in intercept. Random intercepts: participants and items. Model output for time window Sil1. Proportion of looks to the target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "Sil1")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 1685.3

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.0391 -1.0379  0.5436  0.7065  1.2181

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept) 0.000000 0.00000
 id       (Intercept) 0.007493 0.08656
 Residual             0.167802 0.40964
Number of obs: 1529, groups:  item, 64; id, 52

Fixed effects:
               Estimate Std. Error         df t value Pr(>|t|)
(Intercept)   6.408e-01  3.391e-02  2.450e+02  18.897   <2e-16***
mono          6.979e-02  4.811e-02  2.474e+02   1.451   0.1481
PASPF         5.086e-02  4.145e-02  1.474e+03   1.227   0.2200
ACTPF         4.247e-02  4.202e-02  1.476e+03   1.011   0.3123
ACTMF         9.636e-02  4.201e-02  1.475e+03   2.294   0.0219*
mono:PASPF   -5.119e-02  5.890e-02  1.475e+03  -0.869   0.3849
mono:ACTPF   -9.494e-03  5.931e-02  1.476e+03  -0.160   0.8728
mono:ACTMF   -9.455e-02  5.937e-02  1.475e+03  -1.592   0.1115
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.17 Output statistical model Table 35

Treatment contrast for language group and condition. Bilinguals and PAS MF in intercept. Random intercepts: participants and items. Model output for time window Sil2. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "Sil2")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 1622.7

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.9779 -0.9878  0.5891  0.7443  1.1446

Random effects:
 Groups   Name        Variance  Std.Dev.
 item     (Intercept) 4.426e-05 0.006653
 id       (Intercept) 4.171e-03 0.064579
 Residual             1.670e-01 0.408702
Number of obs: 1490, groups:  item, 64; id, 52

Fixed effects:
              Estimate Std. Error         df t value Pr(>|t|)
(Intercept)    0.60142    0.03282  293.53124  18.324   <2e-16***
mono           0.05758    0.04628  350.60490   1.244   0.2143
PASPF          0.09540    0.04225 1430.18519   2.258   0.0241*
ACTPF          0.11030    0.04279  620.89167   2.578   0.0102*
ACTMF          0.10813    0.04261  607.01209   2.538   0.0114*
mono:PASPF    -0.04610    0.06005 1431.44448  -0.768   0.4428
mono:ACTPF    -0.03554    0.06007 1411.37995  -0.592   0.5542
mono:ACTMF    -0.06612    0.06009 1413.06666  -1.100   0.2713
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.18 Output statistical model Table 35

Treatment contrast for language group and condition. Bilinguals and PAS MF in intercept. Random intercepts: participants and items. Model output for time window Sil3. Proportion of looks to the target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + lang_group * cond + (1 | id) + (1 | item)
   Data: subset(d1, win == "Sil3")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 1571.3

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.8020 -1.2463  0.6315  0.8036  1.0690

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept) 0.000000 0.00000
 id       (Intercept) 0.002065 0.04544
 Residual             0.179228 0.42335
Number of obs: 1364, groups:  item, 64; id, 52

Fixed effects:
              Estimate Std. Error         df t value Pr(>|t|)
(Intercept)    0.60166    0.03374  426.79060  17.833   <2e-16***
mono          -0.01349    0.04825  445.36651  -0.280   0.7799
PASPF          0.05176    0.04563 1321.61577   1.134   0.2568
ACTPF          0.09595    0.04651 1316.29886   2.063   0.0393*
ACTMF          0.04692    0.04555 1315.81543   1.030   0.3031
mono:PASPF    -0.01186    0.06521 1320.62603  -0.182   0.8557
mono:ACTPF     0.03780    0.06559 1316.30452   0.576   0.5645
mono:ACTMF     0.04059    0.06483 1317.03681   0.626   0.5313
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.19 Output statistical model Table 37

Repeated contrasts used for condition and language group. Treatment contrast used for accuracy: correct responses in intercept. Random intercepts: participants and items. Random slope: individual adjustment of condition effect for each participant. Proportion of looks to target as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: plt ~ 1 + cond * lang_group * acc + (1+cond|id)+(1|item)
Data: d
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 13985.2

Scaled residuals:
    Min      1Q   Median      3Q     Max
-1.8966 -1.0319   0.5543  0.7977  1.8287

Random effects:
 Groups   Name        Variance Std.Dev.  Corr
Ite m     (Intercept) 0.001406 0.03750
 id       (Intercept) 0.001609 0.04011
          ACTMF-ACTPF 0.011585 0.10764  -0.39
          PASPF-ACTMF 0.013455 0.11600   0.35 -0.49
          PASMF-PASPF 0.005912 0.07689   0.19  0.04 -0.56
 Residual             0.186176 0.43148
Number of obs: 11816, groups:  id, 62; item, 32

Fixed effects:
                Estimate Std. Error         df t value Pr(>|t|)
(Intercept)    6.461e-01  8.493e-03  7.512e+01  76.077   <2e-16***
ACTMF-ACTPF   -1.517e-03  1.931e-02  5.469e+01  -0.079   0.9377
PASPF-ACTMF   -3.553e-02  2.232e-02  7.151e+01  -1.592   0.1159
PASMF-PASPF    1.926e-02  1.645e-02  5.832e+01   1.170   0.2466
bi-mo          2.803e-02  1.414e-02  5.444e+01   1.982   0.0526.
inc.res       -2.728e-01  1.248e-02  8.893e+03 -21.849   <2e-16***
ACTMF-ACTPF:
bi-mo          6.547e-02  3.873e-02  5.497e+01   1.690   0.0966.
PASPF-ACTMF:
bi-mo         -1.235e-02  4.054e-02  5.458e+01  -0.305   0.7618
                Estimate Std. Error         df t value Pr(>|t|)
PASMF-PASPF:
bi-mo         -1.369e-02  3.300e-02  5.845e+01  -0.415   0.6797
ACTMF-ACTPF:
inc.res       -5.426e-02  3.736e-02  6.757e+03  -1.452   0.1464
PASPF-ACTMF:
inc.res        2.499e-02  3.469e-02  6.833e+03   0.720   0.4713
PASMF-PASPF:
inc.res       -2.075e-02  3.352e-02  6.221e+03  -0.619   0.5359
bi-mo:
inc.res       -6.242e-02  2.484e-02  9.260e+03  -2.513   0.0120*
ACTMF-ACTPF:bi-mo:
inc.res       -7.378e-02  7.422e-02  7.365e+03  -0.994   0.3202
PASPF-ACTMF:bi-mo:
inc.res        7.682e-03  6.881e-02  7.123e+03   0.112   0.9111
PASMF-PASPF:bi-mo:
inc.res        1.186e-01  6.631e-02  6.502e+03   1.788   0.0738.
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.20 Output statistical model Table 38

Repeated contrasts for condition. Monolinguals' data.

Model output for time window 1. Random intercepts: participants and items.

Pupil dilation data as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: pup ~ 1 + cond + (1|id) + (1|item)
   Data: subset(dMono, time == "1")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 4496426

Scaled residuals:
     Min      1Q  Median      3Q     Max
 -7.7726 -0.3811  0.0063  0.3836 16.1878

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept)  5189     72.03
 id       (Intercept)  5162     71.85
 Residual             62066    249.13
Number of obs: 324056, groups:  item, 64; id, 26

Fixed effects:
               Estimate Std. Error         df t value Pr(>|t|)
(Intercept)      76.588     16.738     46.125   4.576 3.58e-05 ***
ACTMF-ACTPF     -18.769      1.250 323997.244 -15.015  < 2e-16 ***
PASPF-ACTMF      19.371     18.052     61.758   1.073    0.287
PASMF-PASPF     -49.263      1.267 324018.028 -38.872  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.21 Output statistical model Table 38

Repeated contrasts for condition. Monolinguals' data.

Model output for time window 2. Random intercepts: participants and items.

Pupil dilation data as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: pup ~ 1 + cond + (1|id) + (1|item)
   Data: subset(dMono, time == "2")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 4553254

Scaled residuals:
     Min      1Q  Median      3Q     Max
 -7.3708 -0.3608  0.0129  0.3477 12.3837

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept)  5549     74.49
 id       (Intercept)  6439     80.25
 Residual             72436    269.14
Number of obs: 324538, groups:  item, 64; id, 26

Fixed effects:
               Estimate Std. Error         df t value Pr(>|t|)
(Intercept)     101.759     18.303     43.074   5.560 1.58e-06 ***
ACTMF-ACTPF     -32.354      1.353 324480.164 -23.906  < 2e-16 ***
PASPF-ACTMF       3.842     18.673     61.687   0.206    0.838
PASMF-PASPF       2.979      1.365 324503.549   2.183    0.029 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.22 Output statistical model Table 39
Repeated contrasts for condition. Bilinguals' data.
Model output for time window 1. Random intercepts: participants and items.
Pupil dilation data as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: pup ~ 1 + cond + (1|id) + (1|item)
   Data: subset(dBili, time == "1")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 4193707

Scaled residuals:
    Min      1Q   Median     3Q     Max
-14.7837  -0.4399   0.0438   0.5190  11.1383

Random effects:
 Groups   Name        Variance  Std.Dev.
 item     (Intercept)  1134      33.68
 id       (Intercept)  1385      37.22
 Residual             16427     128.17
Number of obs: 334263, groups:  item, 64; id, 26

Fixed effects:
                Estimate Std. Error        df t value Pr(>|t|)
(Intercept)    4.416e+01  8.434e+00  4.242e+01   5.237  4.8e-06***
ACTMF-ACTPF    1.495e+01  6.694e-01  3.342e+05  22.331  < 2e-16
PASPF-ACTMF    1.657e+00  8.445e+00  6.179e+01   0.196    0.845
PASMF-PASPF   -1.302e+01  6.465e-01  3.342e+05 -20.139  < 2e-16
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.23 Output statistical model Table 39
Repeated contrasts for condition. Bilinguals' data.
Model output for time window 2. Random intercepts: participants and items.
Pupil dilation data as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: pup ~ 1 + cond + (1|id) + (1|item)
   Data: subset(dBili, time == "2")
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 4437575

Scaled residuals:
    Min      1Q    Median      3Q      Max
-16.5805  -0.3598   0.0249   0.3678  24.8797

Random effects:
 Groups   Name        Variance  Std.Dev.
 item     (Intercept)  3315      57.57
 id       (Intercept)  3793      61.59
 Residual             33494     183.01
Number of obs: 334693, groups:  item, 64; id, 26

Fixed effects:
                Estimate  Std. Error        df t value Pr(>|t|)
(Intercept)    6.877e+01  1.407e+01  4.349e+01   4.887 1.43e-05***
ACTMF-ACTPF    2.366e+01  9.501e-01  3.347e+05  24.898  < 2e-16***
PASPF-ACTMF   -5.212e+00  1.442e+01  6.171e+01  -0.361    0.719
PASMF-PASPF   -1.157e+01  9.227e-01  3.347e+05 -12.540  < 2e-16***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.24 Output statistical model Table 40

Repeated contrasts for condition. Monolinguals' data. Treatment contrast used for accuracy: correct responses in the intercept. Random intercepts: participants and items. Pupil dilation data as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: pup ~ 1 + cond * acc + (1|id)+(1|item)
Data: dMono
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 9053474

Scaled residuals:
    Min      1Q  Median      3Q     Max
 -7.7931 -0.3652  0.0032  0.3745 15.5052

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept)  5126     71.60
 id       (Intercept)  5668     75.29
 Residual             67499    259.80
Number of obs: 648594, groups:  item, 64; id, 26

Fixed effects:
              Estimate Std. Error         df t value Pr(>|t|)
(Intercept)     95.812     17.280     44.054   5.545 1.56e-06 ***
ACTMF-ACTPF    -22.512      0.992 648556.946 -22.693  < 2e-16 ***
PASPF-ACTMF     25.779     17.927     61.631   1.438    0.156
PASMF-PASPF    -20.509      1.001 648562.351 -20.484  < 2e-16 ***
inc.res        -46.051      1.120 648507.860 -41.117  < 2e-16 ***
ACTMF-ACTPF:
inc.res        -39.868      3.239 648353.423 -12.311  < 2e-16 ***
PASPF-ACTMF:
inc.res       -133.329      3.134 648326.233 -42.546  < 2e-16 ***
PASMF-PASPF:
inc.res         12.847      3.078 648542.927   4.174 2.99e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.25 Output statistical model Table 41

Repeated contrasts for condition. Bilinguals' data. Treatment contrast used for accuracy: correct responses in the intercept. Random intercepts: participants and items. Pupil dilation data as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: pup ~ 1 + cond * acc + (1|id)+(1|item)
Data: dBili
Control: lmerControl(calc.derivs = FALSE)

REML criterion at convergence: 8687710

Scaled residuals:
    Min      1Q  Median      3Q     Max
-19.8910 -0.3824  0.0368  0.4156 28.9866

Random effects:
 Groups   Name        Variance Std.Dev.
 item     (Intercept)  1811     42.56
 id       (Intercept)  2397     48.96
 Residual             25546    159.83
Number of obs: 668956, groups:  item, 64; id, 26

Fixed effects:
              Estimate Std. Error        df t value Pr(>|t|)
(Intercept)  5.469e+01  1.099e+01 4.101e+01   4.979  1.2e-05 ***
ACTMF-ACTPF  2.244e+01  6.308e-01 6.689e+05  35.583  < 2e-16 ***
PASPF-ACTMF  2.737e+00  1.066e+01 6.166e+01   0.257    0.798
PASMF-PASPF -7.757e+00  6.445e-01 6.689e+05 -12.034  < 2e-16 ***
inc.res      1.943e+01  5.980e-01 6.689e+05  32.499  < 2e-16 ***
ACTMF-ACTPF:
inc.res     -4.102e+01  1.821e+00 6.689e+05 -22.519  < 2e-16 ***
PASPF-ACTMF:
inc.res     -3.114e+01  1.657e+00 6.689e+05 -18.792  < 2e-16 ***
PASMF-PASPF:
inc.res     -1.666e+01  1.508e+00 6.689e+05 -11.053  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.26 Table Degree of balanced bilingualism in the 4-year-old group

Bold values indicate children that are considered balanced bilinguals.

ID	Ratio of German PPVT to Italian	Ratio of German TROG to Italian
1	1.93	**1.13**
2	1.46	**1.04**
3	2.70	1.28
4	1.65	1.97
5	1.90	2.39
6	1.92	1.74
7	1.53	**1.13**
8	1.90	1.30
9	1.34	1.81
10	1.95	NA
11	1.27	1.16
12	NA	NA
13	1.70	1.43
14	1.62	**0.88**
15	1.65	**1.06**
16	1.57	1.33
17	2.73	1.63
18	1.96	1.89
19	2.15	1.67
20	2.13	**1.07**
21	1.49	1.63
22	1.53	1.32
23	1.63	**0.89**
24	2.29	**0.99**
25	1.49	**1.06**
26	1.39	**1.01**

E.27 Output statistical model Table 44

Repeated contrasts used for condition. Language group coded with treatment contrast: monolinguals in intercept. Random intercepts: participants and items. Random slope: individual adjustment of condition effect for each participant. Continuous predictors: flanker effect and global switch cost. Correct responses as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
 Family: binomial  ( logit )
Formula: acc~1+lang_group*cond*(Flanker+GSC)+(1+cond|id)+(1|item)
   Data: d
Control: glmerControl(calc.derivs = FALSE)

     AIC      BIC   logLik deviance df.resid
   739.8    814.8   -353.9    707.8      784

Scaled residuals:
     Min      1Q  Median      3Q     Max
 -3.2598  0.2448  0.3644  0.4851  0.9091

Random effects:
 Groups   Name        Variance Std.Dev.  Corr
 id      (Intercept) 0.20303  0.4506
         PASMF-PASPF 0.26302  0.5129   -1.00
 item    (Intercept) 0.08579  0.2929
Number of obs: 800, groups:  id, 50; item, 32

Fixed effects:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)                2.06709    0.20045  10.312  < 2e-16***
bi                        -0.68941    0.24873  -2.772  0.00558**
PASMF-PASPF               -0.70847    0.35432  -2.000  0.04555*
Flanker                   -0.93910    0.53830  -1.745  0.08106.
GSC                        0.26517    0.22935   1.156  0.24762
bi:PASMF-PASPF            -0.04440    0.44839  -0.099  0.92112
bi:Flanker                 1.04819    0.63014   1.663  0.09623.
bi:GSC                     0.06356    0.34831   0.182  0.85521
PASMF-PASPF:Flanker       -0.40189    0.98125  -0.410  0.68212
PASMF-PASPF:GSC           -0.02913    0.42770  -0.068  0.94569
bi:PASMF-PASPF:Flanker     0.90490    1.13657   0.796  0.42593
bi:PASMF-PASPF:GSC         0.35933    0.63236   0.568  0.56987
```

E.28 Output statistical model Table 45

Repeated contrasts used for condition. Language group coded with treatment contrast: bilinguals in intercept. Random intercepts: participants and items. Random slope: individual adjustment of condition effect for each participant. Continuous predictors: flanker effect and global switch cost. Correct responses as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
 Family: binomial  ( logit )
Formula: acc~1+lang_group*cond*(Flanker+GSC)+(1+cond|id)+(1|item)
   Data: d
Control: glmerControl(calc.derivs = FALSE)

     AIC      BIC   logLik deviance df.resid
   739.8    814.8   -353.9    707.8      784

Scaled residuals:
     Min      1Q  Median      3Q     Max
 -3.2617  0.2450  0.3644  0.4850  0.9088

Random effects:
 Groups   Name        Variance Std.Dev.   Corr
 id      (Intercept) 0.20330  0.4509
        PASMF-PASPF 0.26311  0.5129   -1.00
 item   (Intercept) 0.08571  0.2928
Number of obs: 800, groups:  id, 50; item, 32

Fixed effects:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)                1.37757    0.16457   8.371  < 2e-16***
mo                         0.69004    0.24880   2.774  0.00555
PASMF-PASPF               -0.75186    0.27483  -2.736  0.00623**
Flanker                    0.10971    0.32762   0.335  0.73773
GSC                        0.32864    0.26185   1.255  0.20946
mo:PASMF-PASPF             0.04397    0.44844   0.098  0.92190
mo:Flanker                -1.04866    0.63030  -1.664  0.09616.
mo:GSC                    -0.06250    0.34844  -0.179  0.85766
PASMF_PASPF:Flanker        0.50160    0.57903   0.866  0.38634
PASMF-PASPF:GSC            0.32858    0.46721   0.703  0.48188
mo:PASMF-PASPF:Flanker    -0.90465    1.13667  -0.796  0.42610
mo:PASMF-PASPF:GSC        -0.35824    0.63250  -0.566  0.57113
```

E.29 Output statistical model Table 47

Repeated contrasts used for condition. Only the bilingual group is included here. Random intercepts: participants and items. Random slope: individual adjustment of condition effect for each participant. Continuous predictors in the model: language experience, number of older siblings and parents' education. Correct responses as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
 Family: binomial  ( logit )
 Formula:
 acc~1+cond*(lang_exp+nr_older_sib+L_edu)+(1+cond|id)+(1|item)
 Data: df
Control: glmerControl(calc.derivs = FALSE)

     AIC      BIC   logLik deviance df.resid
   335.6    380.8   -155.8    311.6      308

Scaled residuals:
     Min      1Q   Median     3Q     Max
 -3.1062  0.2907   0.4037 0.5783  0.7728

Random effects:
 Groups   Name         Variance  Std.Dev. Corr
 item   (Intercept)  6.175e-08 0.0002485
 id     (Intercept)  1.837e-01 0.4285785
       PASMF-PASPF  3.237e-03 0.0568962 1.00
Number of obs: 336, groups:  item, 32; id, 21

Fixed effects:
                         Estimate Std. Error z value Pr(>|z|)
(Intercept)               1.53533    0.18735   8.195 2.51e-16 ***
PASMF-PASPF              -0.87277    0.31852  -2.740  0.00614 **
lang_exp                 -0.02095    0.01279  -1.637  0.10153
no_sibling               -0.05040    0.34409  -0.146  0.88354
L_edu                    -0.11965    0.28485  -0.420  0.67444
PASMF-PASPF:lang_exp      0.03316    0.02084   1.591  0.11161
PASMF-PASPF:L_edu         0.24133    0.58311   0.414  0.67897
PASMF-PASPF:no_sibling    0.09485    0.48277   0.196  0.84425
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.30 Output statistical model Table 48

Repeated contrasts used for condition. Only the monolingual group is included here. Random intercept: participants and items. Random slope: individual adjustment of condition effect for each participant. Continuous predictors: number of older siblings and parents' education. Correct responses as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
 Family: binomial  ( logit )
Formula: acc~1+cond*(L_edu+nr_older_sib)+(1+cond|id)+(1|item)
 Data: df
Control: glmerControl(calc.derivs = FALSE)

     AIC      BIC   logLik deviance df.resid
   189.7    226.3    -84.8    169.7      278

Scaled residuals:
     Min      1Q  Median     3Q      Max
 -4.1826  0.1245  0.2337 0.3271  1.0337

Random effects:
 Groups   Name       Variance  Std.Dev.  Corr
 item    (Intercept)  0.6767   0.8226
 id      (Intercept)  0.2297   0.4792
         PASMF-PASPF  2.4128   1.5533   -0.79
Number of obs: 288, groups:  item, 32; id, 18

Fixed effects:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)              6.3866  5207.1611   0.001    0.999
PASMF-PASPF             -7.8228 10414.3223  -0.001    0.999
L_edu                    0.3125     0.3675   0.850    0.395
no_sibling              10.0108 14133.7231   0.001    0.999
PASMF-PASPF:L_edu        0.2142     0.8360   0.256    0.798
PASMF-PASPF:no_sibling -18.1740 28267.4462  -0.001    0.999
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.31 Output statistical model Table 50

Repeated contrasts used for condition. Language group coded with treatment contrast: bilinguals in the intercept. Random intercepts: participants and items. Random slope: individual adjustment of condition effect for each participant. Continuous predictors: language knowledge, memory, and reasoning ability. Correct responses as dependent dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Family: binomial  ( logit )
Formula:
acc~1+lang_group*cond*(lang_knowl+mem+reas)+(1+cond|id)+(1|item)
Data: d
Control: glmerControl(calc.derivs = FALSE)

    AIC      BIC   logLik deviance df.resid
  745.0    839.0   -352.5    705.0      796

Scaled residuals:
     Min      1Q    Median      3Q     Max
 -4.1237  0.2155    0.3431  0.5181  0.9161

Random effects:
 Groups   Name       Variance Std.Dev. Corr
 id     (Intercept) 0.09966  0.3157
        PASMF-PASPF 0.03964  0.1991 -1.00
 item   (Intercept) 0.10860  0.3295
Number of obs: 816, groups:  id, 51; item, 32

Fixed effects:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)               1.535788   0.196170   7.829 4.92e-15***
mo                        0.491239   0.284963   1.724  0.08473.
PASMF-PASP               -0.911043   0.346617  -2.628  0.00858**
lang_knowl                0.025760   0.017198   1.498  0.13417
mem                       0.047209   0.016730   2.822  0.00477
reas                     -0.038665   0.021516  -1.797  0.07233.
mo:PASMF-PASPF            0.241767   0.532590   0.454  0.64987
mo:lang_knowl             0.006998   0.030398   0.230  0.81791
mo:mem                   -0.022756   0.029368  -0.775  0.43842
mo:reas                   0.023447   0.031031   0.756  0.44990
```

```
                          Estimate Std. Error z value Pr(>|z|)
PASMF-PASPF:lang_knowl    -0.004430   0.030859  -0.144  0.88584
PASMF-PASPF:mem           -0.068892   0.031408  -2.193  0.02827*
PASMF-PASPF:reas           0.064604   0.038947   1.659  0.09716.
mo:PASMF-PASPF:lang_knowl  0.113064   0.056667   1.995  0.04602*
mo:PASMF-PASPF:mem         0.018967   0.055157   0.344  0.73094
mo:PASMF-PASPF:reas        0.002012   0.057543   0.035  0.97210
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E.32 Output statistical model Table 51

Repeated contrasts used for condition. Language group coded with treatment contrast: monolinguals in the intercept. Random intercepts: participants and items. Random slope: individual adjustment of condition effect for each participant. Continuous predictors: language knowledge, memory, and reasoning ability. Correct responses as dependent variable.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Family: binomial  ( logit )
Formula:
acc~1+lang_group*cond*(lang_knowl+mem+reas)+(1+cond|id)+(1|item)
Data: d
Control: glmerControl(calc.derivs = FALSE)

    AIC      BIC   logLik deviance df.resid
  745.0    839.0   -352.5    705.0      796

Scaled residuals:
    Min      1Q    Median      3Q     Max
-4.1284  0.2154    0.3431  0.5186  0.9164

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 0.09891  0.3145
          PASMF-PASPF 0.03952  0.1988 -1.00
 item     (Intercept) 0.10841  0.3293
Number of obs: 816, groups:  id, 51; item, 32

Fixed effects:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)               2.026736   0.222123   9.124   <2e-16***
bi                       -0.490693   0.284772  -1.723   0.0849.
PASMF-PASPF              -0.665675   0.402706  -1.653   0.0983.
lang_knowl                0.032794   0.025068   1.308   0.1908
mem                       0.024501   0.024120   1.016   0.3097
reas                     -0.015195   0.022417  -0.678   0.4979
bi:PASMF-PASPF           -0.245913   0.532514  -0.462   0.6442
bi:lang_knowl            -0.007005   0.030376  -0.231   0.8176
bi:mem                    0.022663   0.029344   0.772   0.4399
bi:reas                  -0.023556   0.031009  -0.760   0.4475
```

```
                           Estimate Std. Error z value Pr(>|z|)
PASMF-PASPF:lang_knowl     0.108474   0.047404   2.288   0.0221*
PASMF-PASPF:mem           -0.049794   0.045282  -1.100   0.2715
PASMF-PASPF:reas           0.066840   0.042563   1.570   0.1163
mo:PASMF-PASPF:lang_knowl -0.112876   0.056656  -1.992   0.0463*
mo:PASMF-PASPF:mem        -0.018926   0.055138  -0.343   0.7314
mo:PASMF-PASPF:reas       -0.002329   0.057535  -0.040   0.9677
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

References

Abbot-Smith, K., & Behrens, H. (2006). How Known Constructions Influence the Acquisition of Other Constructions: The German Passive and Future Constructions. *Cognitive Science*, *30*, 995–1026. https://doi.org/10.1207/s15516709cog0000_61

Abbot-Smith, K., Chang, F., Rowland, C., Ferguson, H., & Pine, J. (2017). Do two and three year old children use an incremental first-NP-as-agent bias to process active transitive and passive sentences? A permutation analysis. *PLoS ONE*, *12*(10), 1–20. https://doi.org/10.31234/osf.io/fmkcj

Abbot-Smith, K., & Serratrice, L. (2015). Word order, referential expression, and case cues to the acquisition of transitive sentences in Italian. *Journal of Child Language*, *42*, 1–31. https://doi.org/10.1017/S0305000913000421

Anderson, J. A. E., Hawrylewicz, K., & Bialystok, E. (2018). Who is bilingual? Snapshots across the lifespan . *Bilingualism: Language and Cognition*, 1–12. https://doi.org/10.1017/S1366728918000950

Argyri, E., & Sorace, A. (2007). Crosslinguistic influence and language dominance in older bilingual children. *Bilingualism: Language and Cognition*, *10*(1), 79–99. https://doi.org/10.1017/S1366728906002835

Armon-Lotem, S., Haman, E., Jensen de López, K., Smoczynska, M., Yatsushiro, K., Szczerbinski, M., van Hout, A., Dabašinskienė, I., Gavarró, A., Hobbs, E., Kamandulytė-Merfeldienė, L., Katsos, N., Kunnari, S., Nitsiou, C., Sundahl Olsen, L., Parramon, X., Sauerland, U., Torn-Leesik, R., & van der Lely, H. (2016). A large-scale cross-linguistic investigation of the acquisition of passive. *Language Acquisition*, *23*(1), 27–56. https://doi.org/10.1080/10489223.2015.1047095

Armon-Lotem, S., Walters, J., & Gagarina, N. (2011). The impact of internal and external factors on linguistic performance in the home language and in L2 among Russian-Hebrew and Russian-German preschool children. *Linguistic Approaches to Bilingualism*, *1*(3), 291–317. https://doi.org/10.1075/lab.1.3.04arm

Arosio, F., & Giustolisi, B. (2019). Agreement and Interference in Direct Object Clitic Production in Italian Monolingual Children. *Frontiers in Communication*, *3*(66), 1–23. https://doi.org/10.3389/fcomm.2018.00066

Arosio, F., Guasti, M. T., & Stucchi, N. (2010). Disambiguating Information and Memory Resources in Children’s Processing of Italian Relative Clauses. *Journal of Psycholinguistic Research*, *40*(2), 137–154. https://doi.org/10.1007/s10936-010-9160-0

Arriaga, R. I., Fenson, L., Cronan, T., & Pethick, S. J. (1998). Scores on the MacArthur Communicative Development Inventory of children from low-and middle-income families. *Applied Psycholinguistics*, *19*, 209–223. https://www.cambridge.org/core.

Arsenian, S. (1967). *Bilingualism and mental development*. Columbia University Press.

Aschermann, E., Gülzow, I., & Wendt, D. (2004). Differences in the Comprehension of Passive Voice in German- and English-Speaking Children. *Swiss Journal of Psychology*, *63*(4), 235–245. https://doi.org/10.1024/1421-0185.63.4.235

Baayen, H., Davidson, D. J., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bader, M., Meng, M., & Bayer, J. (2000). Case and Reanalysis. *Journal of Psycholinguistic Research*, *29*(1), 37–52.

Bahlmann, J., Rodriguez-Fornells, A., Rotte, M., & Münte, T. F. (2007). An fMRI study of canonical and noncanonical word order in German. *Human Brain Mapping*, *28*(10), 940–949. https://doi.org/10.1002/hbm.20318

Bain, B. (1974). Bilingualism and cognition: Toward a general theory. In S. T. Carey (Ed.), *Bilingualism, biculturalism, and education.* (pp. 119–128). Canada: University of Alberta.

Barac, R., & Bialystok, E. (2011). Cognitive development of bilingual children. *Language Teaching*, *44*(1), 36–54. https://doi.org/10.1017/S0261444810000339

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*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, E. (1976). *Language and context: The acquisition of pragmatics*. Academic Press.

Bates, E., MacWhinney, B., Caselli, C., Devescovi, A., Natale, F., & Venza, V. (1984). A Cross-Linguistic Study of the Development of Sentence Interpretation Strategies. *Child Development*, *55*, 341–354.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292. https://doi.org/10.1037//0033-2909.91.2.276

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (Second, pp. 142–162). Cambridge University Press.

Bedore, L. M., Pena, E. D., Summers, C. L., Boerger, K. M., Resendiz, M. D., Greene, K., Bohman, T. M., & Gillam, R. B. (2012). The measure matters: Language dominance profiles across measures in Spanish-English bilingual children. *Bilingualism: Language and Cognition*, *15*(3), 616–629. https://doi.org/10.1017/S1366728912000090

Belletti, A. (1999). Italian/Romance Clitics: Structure and Derivation. In H. van Riemsdijk (Ed.), *Clitics in the Languages of Europe* (pp. 543–579). Mouton De Gruyter.

Belletti, A., & Guasti, M. T. (2015). *The acquisition of Italian. Morphisyntax and its interfaces in different modes of acquisition.* John Benjamins Publishing Company.

Belletti, A., & Rizzi, L. (1988). Psych-Verbs and θ-Theory. *Natural Language and Linguistic Theory*, *6*, 291–352. https://www.jstor.org/stable/4047649

Benincà, P. (1988). L'ordine degli elementi della frase e le costruzioni marcate. In L. Renzi & G. Salvi (Eds.), *Grande Grammatica Italiana di Consultazione* (Vol. 1, pp. 129–194). Il Mulino.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–352). Wiley. https://doi.org/10.1093/acprof

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development*, *70*(3), 636–644. https://doi.org/10.1111/1467-8624.00046

Bialystok, E. (2005). Consequences of Bilingualism for Cognitive Development. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 417–432). Oxford University Press.

Bialystok, E. (2021). Bilingualism as a Slice of Swiss Cheese. *Frontiers in Psychology*, *12*(769323), 1–6. https://doi.org/10.3389/fpsyg.2021.769323

Bialystok, E., & Barac, R. (2012). Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition*, *122*, 67–73.

https://doi.org/10.1016/j.cognition.2011.08.003

Bialystok, E., Klein, R., Craik, F. I. M., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging*, *19*(2), 290–303. https://doi.org/10.1037/0882-7974.19.2.290

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2009). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*(4), 525–531. https://doi.org/10.1017/S1366728909990423

Bialystok, E., & Majumder, S. (1998). The relationship between bilingualism and the development of cognitive processes in problem solving. *Applied Psycholinguists*, *19*(1), 69–85. https://doi.org/10.1017/S0142716400010584

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science*, *7*(3), 325–339. https://doi.org/10.1111/j.1467-7687.2004.00351.x

Biran, M., & Ruigendijk, E. (2015). Do case and gender information assist sentence comprehension and repetition for German- and Hebrew-speaking children? *Lingua*, *164*, 215–238. https://doi.org/10.1016/j.lingua.2015.06.012

Bishop, D. V. M. (2003). *Test for Reception of Grammar (TROG-2).* Pearson Assessment.

Blom, E., & Bosma, E. (2016). The sooner the better? An investigation into the role of age of onset and its relation with transfer and exposure in bilingual Frisian-Dutch children. *Journal of Child Language*, *43*, 581–607. https://doi.org/10.1017/S0305000915000574

Blom, E., Küntay, A. C., Messer, M., Verhagen, J., & Leseman, P. (2014). The benefits of being bilingual: Working memory in bilingual Turkish–Dutch children. *Journal of Experimental Child Psychology*, *128*, 105–119. https://doi.org/10.1016/j.jecp.2014.06.007

Blom, E., & Vasić, N. (2011). The production and processing of determiner–noun agreement in child L2 Dutch. *Linguistic Approaches to Bilingualism*, *3*(1), 265–290. https://doi.org/10.1075/lab.1.3.03blo

Bock, J. K. (1986). Syntactic Persistence in Language Production. *Cognitive Psychology*, *18*, 355–387.

Bohman, T. M., Bedore, L. M., Peña, E. D., Mendez-Perez, A., & Gillam, R. B. (2010). What you hear and what you say: language performance in Spanish–English bilinguals. *International Journal of Bilingual Education and Bilingualism*, *13*(3), 325–344. https://doi.org/10.1080/13670050903342019

Bohnacker, U. (2007). *Will they keep on starting their sentences the German way? The prefield in advanced learners of Swedish*. Manuscript submitted for publication.

Bohnacker, U., & Rosén, C. (2008). The Clause-Initial Position in L2 German declaratives. Transfer of Information Structure. *SSLA*, *30*, 511–538. https://doi.org/10.1017/S0272263108080741

Bonfieni, M., Branigan, H. P., Pickering, M. J., & Sorace, A. (2020). Cognitive control in bilinguals: Effects of language experience and individual variability. *Bilingualism: Language and Cognition*, *23*, 219–230. https://doi.org/10.1017/S1366728918001086

Borer, H., & Wexler, K. (1987). The maturation of syntax. In T. Roeper & E. Williams (Eds.), *Parameter Setting* (pp. 123–172). Reidel Publishing Company.

Brandt-Kobele, O. C., & Höhle, B. (2010). What asymmetries within comprehension reveal about asymmetries between comprehension and production: The case of verb inflection in language acquisition. *Lingua*, *120*(8), 1910–1925. https://doi.org/10.1016/j.lingua.2010.02.008

Brandt, S., Lieven, E., & Tomasello, M. (2016). German Children's Use of Word Order and Case

Marking to Interpret Simple and Complex Sentences: Testing Differences Between Constructions and Lexical Items. *Language Learning and Development*, *12*(2), 156–182. https://doi.org/10.1080/15475441.2015.1052448

Bridges, K., & Hoff, E. (2014). Older sibling influences on the language environment and language development of toddlers in bilingual homes. *Applied Psycholinguistics*, *35*, 225–241. https://doi.org/10.1017/S0142716412000379

Broekhuis, H. (2007). Object shift and subject shift. *The Journal of Comparative Germanic Linguistics*, *10*, 109–141. https://doi.org/10.1007/s10828-007-9009-6

Bulheller, S., & Häcker, H. (2002). *Coloured Progressive Matrices*. Pearson Assessment.

Burmester, J., Spalek, K., & Wartenburger, I. (2014). Context updating during sentence comprehension: The effect of aboutness topic. *Brain and Language*, *137*, 62–76. https://doi.org/10.1016/j.bandl.2014.08.001

Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition*, *130*(3), 278–288. https://doi.org/10.1016/j.cognition.2013.11.015

Cantone, K., Kupisch, T., Müller, N., & Schmitz, K. (2006). *Rethinking Language Dominance in Bilingual Children* (pp. 307–343).

Cardinaletti, A., & Starke, M. (1999). The typology of structural deficiency: A case study of the three classes of pronouns. In H. van Riemsdijk (Ed.), *Clitics in the Languages of Europe* (pp. 145–233). Mouton De Gruyter.

Carnie, A. (2007). *Syntax: A generative Introduction* (Second). Blackwell Publishing Ltd.

Carroll, J. B. (1962). Prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training research and education*. University of Pittsburgh Press.

Cepeda, N. J., Kramer, A. F., & de Santher, J. C. M. G. (2001). Changes in Executive Control Across the Life Span: Examination of Task-Switching Performance. *Developmental Psychology*, *37*(5), 715–730.

Chiat, S., & Polišenská, K. (2016). A Framework for Crosslinguistic Nonword Repetition Tests: Effects of Bilingualism and Socioeconomic Status on Children's Performance. *Journal of Speech, Language, and Hearing Research*, *59*, 1179–1189. https://doi.org/10.1044/2016_JSLHR-L-15-0293

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*, 41–61. https://doi.org/10.1016/j.jecp.2010.01.003

Chomsky, N. (1981). *Lectures on government and binding.* Foris Publications Holland.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin and use.* Praeger.

Chomsky, N. (1995). *The Minimalist Program*. The MIT Press.

Chondrogianni, Vasiliki, & Marinis, T. (2012). Production and processing asymmetries in the acquisition of tense morphology by sequential bilingual children. *Bilingualism: Language and Cognition*, *15*(1), 5–21. https://doi.org/10.1017/S1366728911000368

Chondrogianni, Vasiliki, Vasić, N., Marinis, T., & Blom, E. (2015). Production and on-line comprehension of definiteness in English and Dutch by monolingual and sequential bilingual children. *Second Language Research*, *31*(3), 309–341. https://doi.org/10.1177/0267658314564461

Chondrogianni, Vicky, & Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, tense morphology and morpho-syntax in successive bilingual children. *Linguistic Approaches to Bilingualism*, *1*(3), 318–345. https://doi.org/10.1075/lab.1.3.05cho

Christoffels, I. K., Kroll, J. F., & Bajo, M. T. (2013). Introduction to Bilingualism and Cognitive Control. *Frontiers in Psychology*, *4*(199), 1–3. https://doi.org/10.3389/fpsyg.2013.00199

Cinque, G. (1990). *Types of A' Dependencies*. MIT Press.

Clahsen, H., Eisenbeiß, S., & Penke, M. (1996). Lexical learning in early syntactic development. In H. Clahsen & W. Rutherford (Eds.), *Generative Perspectives on Language Acquisition* (pp. 129–159). Benjamins.

Clahsen, Harald, & Felser, C. (2006). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, *27*, 107–126. https://doi.org/10.1017.S0142716406060206

Cohen, A. L. (2013). Software for the automatic correction of recorded eye fixation locations in reading experiments. *Behavior Research Methods*, *45*(3), 679–683. https://doi.org/10.3758/s13428-012-0280-3

Cornips, L., & Hulk, A. (2008). Factors of success and failure in the acquisition of grammatical gender in Dutch. *Second Language Research*, *24*(3), 267–295. https://doi.org/10.1177/0267658308090182

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition*, *113*(2), 135–149. https://doi.org/10.1016/j.cognition.2009.08.001

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, *106*, 59–86. https://doi.org/10.1016/j.cognition.2006.12.013

Cristante, V. (2016). *The Processing of Non-Canonical Sentences in Children with German as a First or Second Language and German Adults.* [Doctoral dissertation. Westfälischen Wilhelms-Universität Münster]. https://miami.uni-muenster.de/Record/64334bda-5675-4d0e-9f4d-17c3de57ac8a

Cristante, V., & Schimke, S. (2018). Verarbeitung von Passiv- und OVS-Sätzen bei Kindern mit Deutsch als früher L2. In S. Schimke & H. Hopp (Eds.), *Sprachverarbeitung im Zweitspracherwerb* (pp. 169–192). De Gruyter Mouton.

Cristante, V., & Schimke, S. (2020). The processing of passive sentences in German: Evidence from an eye-tracking study with seven- and ten-year-olds and adults. *Language, Interaction and Acquisition*, *11*(2), 163–195. https://doi.org/10.1075/lia.19013.cri

Cruschina, S. (2010). Syntactic Extraposition and clitic resumption in Italian. *Lingua*, *120*, 50–73. https://doi.org/10.1016/j.lingua.2009.04.002

Cummins, J. (1976). The influence of bilingualism on cognitive growth: A synthesis of research findings and explanatory hypotheses. *Working Papers on Bilingualism*, *9*, 1–43.

Darcy, N. T. (1953). A review of the literature on the effects of bilingualism upon the measurement of intelligence. *Journal of Genetic Psychology*, *82*, 21–57.

Darcy, N. T. (1963). Bilingualism and the measurement of intelligence: A review of a decade of research. *Journal of Genetic Psychology*, *103*, 259–282.

de Bruin, A. (2019). Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioral Sciences*, *9*(3). https://doi.org/10.3390/bs9030033 de Bruin, A., Treccani, B., & Sala, S. Della. (2015). Cognitive Advantage in Bilingualism: An Example of Publication Bias ? *Psychological Science*, *26*(1), 99–107. https://doi.org/10.1177/0956797614557866

De Cat, C., Gusnanto, A., & Serratrice, L. (2018). Identifying a threshold for the executive function advantage in bilingual children. *Studies in Second Language Acquisition*, *40*, 119–151. https://doi.org/10.1017/S0272263116000486

De Houwer, A. (1990). *The acquisition of two languages from birth: A case study*. Cambridge University Press.

De Houwer, A. (2007). Parental language input patterns and children's bilingual use. *Applied Psycholinguistics*, *28*(3), 411–424. https://doi.org/10.1017.S0142716407070221

De Houwer, A. (2014). The absolute frequency of maternal input to bilingual and monolingual children: A first comparison. In T. Grüter & J. Paradis (Eds.), *Input and Experience in Bilingual Development* (pp. 37–58). John Benjamins Publishing Company. https://doi.org/10.1075/tilar.13.03deh

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

Dittmar, M., Abbot-Smith, K., & Lieven, E. (2014). Familiar Verbs are not always easier than Novel Verbs: How German Pre-School Children Comprehend Active and Passive Sentences. *Cognitive Science*, *38*, 128–151. https://doi.org/10.1111/cogs.12066

Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2008). German Children's Comprehension of Word Order and Case Marking in Causative Sentences. *Child Development*, *79*(4), 1152–1167. https://doi.org/https://doi.org/10.1111/j.1467-8624.2008.01181.x

Döpke, S. (1998). Competing language structures: the acquisition of verb placement by bilingual German-English children. *Journal of Child Language*, *25*, 555–584. https://doi.org/10.1017/S0305000998003584

Duñabeitia, J. A., Andrés Hernández, J., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2014). The Inhibitory Advantage in Bilingual Children Revisited: Myth or Reality? *Experimental Psychology2*, *61*(3), 234–251. https://doi.org/10.1027/1618-3169/a000243

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test, Fourth Edition*. Pearson Education.

Ehrenhofer, L. (2018). *Argument Roles in Adult and Child*. [Doctoral dissertation. University of Maryland].

Engel, U. (1974). Syntaktische Besonderheiten der deutschen Alltagssprache. In G. Ungeheuer (Ed.), *Gesprochene Sprache, Jahrbuch 1972 (Sprache der Gegenwart 26)* (pp. 199–228). Pädagogischer Verlag Schwann.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149.

Feng, S., Legault, J., & Yang, L. (2015). Differences in grammatical processing strategies for active and passive sentences: An fMRI study. *Journal of Neurolinguistics*, *33*, 104–117. https://doi.org/10.1016/j.jneuroling.2014.09.002

Fenson, L., Dale, P. S., Reznick, J. S., Thai, D., Bates, E., Hartung, J. P., Pethick, S., & Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. Singular Publishing Group.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*, 164–

203. https://doi.org/10.1016/S0010-0285(03)00005-7

Ferreira, F., & Henderson, J. M. (1998). Syntactic Reanalysis, Thematic Processing, and Sentence Comprehension. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 73–100). Kluwer Academic Publishers.

Fillmore, C. J. (1968). The Case for case. In E. Bach & R. T. Harms (Eds.), *Universals in Linguistics Theory* (pp. 1–134). Holt, Rhehart and Winston.

Fisher, C., Church, B. A., & Chambers, K. E. (2004). Learning to identify spoken words. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a Lexicon* (pp. 3–40). The MIT Press.

Fox, D., & Grodzinsky, Y. (1998). Children's Passive: A View from the By-Phrase. *Linguistics Inquiry*, *29*(2), 311–332.

Fox, A. V. (2009). *TROG-D*. Schulz-Kirchner Verlag GmbH.

Frank, M. C., Vul, E., & Saxe, R. (2012). Measuring the Development of Social Attention Using Free-Viewing. *Infancy*, *17*(4), 355–375. https://doi.org/10.1111/j.1532-7078.2011.00086.x

Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies*. Indiana University Linguistics Club.

Frey, W. (2006). Contrast and movement to the German prefield. In V. Molnár & S. Winkler (Eds.), *The architecture of focus* (pp. 235–264). De Gruyter Mouton.

Friedman, N. P. (2019). Research on individual differences in executive functions: Implications for the bilingual advantage hypothesis. In I. Sekerina, L. Spradlin, & V. Valian (Eds.), *Bilingualism, Executive Function, and Beyond: Questions and insights* (pp. 209–222). John Benjamins Publishing Company.

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

Friedmann, N., Rizzi, L., & Belletti, A. (2017). No case for Case in locality: Case does not help interpretation when intervention blocks A-bar chains. *Glossa: A Journal of General Linguistics*, *2(1)*(33), 1–18. https://doi.org/10.5334/gjgl.165

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, *10*, 483–527. https://doi.org/10.1016/0885-2014(95)90024-1

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). MAIN: Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics*, *56*, 139–142.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory from 4 to 15 Years of Age. *Developmental Psychology*, *40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Gathercole, V. C. M. (2016). Factors Moderating Proficiency in Bilingual Speakers. In E. Nicoladis & S. Montanari (Eds.), *Bilingualism across the lifespan: Factors moderating language proficiency* (pp. 123–140). American Psychological Association. https://doi.org/10.1515/9783110341249-009

Gathercole, V. C. M., Kennedy, I., & Thomas, E. I. (2016). Socioeconomic level and bilinguals' performance on language and cognitive measures . *Bilingualism: Language and Cognition*, *19*(5), 1057–1078. https://doi.org/10.1017/S1366728915000504

Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Guasch, N. V., Roberts, E. J., Hughes, E. K., & Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Frontiers in Psychology*, *5*(FEB), 11. https://doi.org/10.3389/FPSYG.2014.00011/BIBTEX

Genesee, F. (1989). Early bilingual development: one language or two? *Journal of Child Language*, *16*, 161–179. https://doi.org/10.1017/S0305000900013490

Genesee, F., Nicoladis, E., & Paradis, J. (1995). Language differentiation in early bilingual development. *Journal of Child Language*, *22*, 611–631. https://doi.org/10.1017/S0305000900009971

Genesee, F., Paradis, J., & Crago, M. (2004). *Dual Language Development and Disorders: A Handbook on Bilingualism and Second Language Learning.* Brookes Publishing.

Giovannoli, J., Martella, D., Federico, F., Pirchio, S., & Casagrande, M. (2020). The Impact of Bilingualism on Executive Functions in Children and Adolescents: A Systematic Review Based on the PRISMA Method. *Frontiers in Psychology*, *11*, 1–29. https://doi.org/10.3389/fpsyg.2020.574789

Golberg, H., Paradis, J., & Crago, M. (2008). Lexical acquisition over time in minority first language children learning English as a second language. *Applied Psycholinguistics*, *29*, 41–65. https://doi.org/10.1017/S014271640808003X

Gold, B. T., Kim, C., Johnson, N. F., Kryscio, R. J., & Smith, C. D. (2013). Lifelong Bilingualism Maintains Neural Efficiency for Cognitive Control in Aging. *The Journal of Neuroscience*, *33*(2), 387–396. https://doi.org/10.1523/JNEUROSCI.3837-12.2013

Grewendorf, G. (2002). *Minimalistische Syntax*. Francke.

Grosjean, F. (1998). Transfer and language mode. *Bilingualism: Language and Cognition*, *1*, 175–176.

Gruber, J. (1976). *Lexical Structures in Syntax and Semantics*. North-Holland Publishing Company.

Guasti, M. T. (1993). Verb Syntax in Italian Child Grammar: Finite and Nonfinite Verbs. *Language Acquisition*, *3*(1), 1–40. https://doi.org/10.1207/s15327817la0301_1

Haegeman, L. (1994). *Introduction to Government and Binding Theory* (Second edi). Blackwell Publishers Ltd.

Hakuta, K., & Diaz, R. M. (1985). The Relationship Between Degree of Bilingualism and Cognitive Ability: A Critical Discussion and Some New Longitudinal Data. In K. E. Nelson (Ed.), *Children's language* (Vol. 5, pp. 319–344). Lawrence Erlbaum Associates, Inc.

Haman, E., Wodniecka, Z., Marecka, M., Szewczyk, J., Białecka-Pikul, M., Otwinowska, A., Mieszkowska, K., Łuniewska, M., Kołak, J., Mi, A., Kacprzak, A., Banasik, N., & Fory, M. (2017). How Does L1 and L2 Exposure Impact L1 Performance in Bilingual Children? Evidence from Polish-English Migrants to the United Kingdom. *Frontiers in Psychology*, *8*(1444), 1–21. https://doi.org/10.3389/fpsyg.2017.01444

Hanne, S., Burchert, F., & De Bleser, R. (2015). Sentence comprehension and morphological cues in aphasia: What eye-tracking reveals about integration and prediction. *Journal of Neurolinguistics*, *34*, 83–111. https://doi.org/10.1016/j.jneuroling.2014.12.003

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experiences of young American children*. Paul H. Brookes Publishing.

Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in spanish-english bilinguals. *Psychological Science*, *15*(6), 409–414. https://doi.org/10.1111/j.0956-7976.2004.00693.x

Hervais-Adelman, A., Moser-Mercer, B., Michel, C. M., & Golestani, N. (2015). fMRI of Simultaneous Interpretation Reveals the Neural Basis of Extreme Language Control. *Cerebral Cortex*, *25*, 4727–479. https://doi.org/10.1093/cercor/bhu158

Hoff, E., & Core, C. (2013). Input and language development in bilingually developing children. *Seminars in Speech and Language*, *34*(4), 215–226. https://doi.org/10.1055/s-0033-1353448

Hoff, E., Laursen, B., & Tardif, T. (2002). Socioeconomic status and parenting. In M. H. Bornstein (Ed.), *Handbook of parenting. Ecology and Biology of Parenting: Vol. II* (pp. 161–188). Lawrence Erlbaum.

Hoff, E., Welsh, S., Place, S., & Ribot, K. M. (2014). Properties of dual language input that shape bilingual development and properties of environments that shape dual language input. In T. Grüter & J. Paradis (Eds.), *Input and Experience in Bilingual Development* (pp. 119–140). John Benjamins Publishing Company. https://doi.org/10.1075/tilar.13.07hof

Höhle, B., Fritzsche, T., & Muller, A. (2016). Children's comprehension of sentences with focus particles and the role of cognitive control: An eye tracking study with German- learning 4-year-olds. *PLoS ONE*, *11*(3), 1–27. https://doi.org/10.1371/journal.pone.0149870

Horgan, D. (1978). The development of the full passive. *Journal of Child Language*, *5*(1), 65–80. https://doi.org/10.1017/S030500090000194X

Hsin, L., Legendre, G., & Omaki, A. (2013). Priming cross-linguistic Interference in Spanish-English bilingual children . In S. Baiz, N. Goldman, & R. Hawkes (Eds.), *Proceedings of the 37th annual Boston University Conference on Language Development* (pp. 165–177). Cascadilla Press. https://www.researchgate.net/publication/269305384

Huang, Y. T., Leech, K., & Rowe, M. L. (2017). Exploring socioeconomic differences in syntactic development through the lens of real-time processing. *Cognition*, *159*, 61–75. https://doi.org/https://doi.org/10.1016/j.cognition.2016.11.004

Huang, Y. T., Zheng, X., Meng, X., & Snedeker, J. (2013). Children's assignment of grammatical roles in the online processing of Mandarin passive sentences. *Journal of Memory and Language*, *69*(4), 589–606. https://doi.org/10.1016/j.jml.2013.08.002

Hulk, A. (2000). Non-Selective Access and Activation in Child Bilingualism: The Syntax. In S. Döpke (Ed.), *Cross-linguistic Structures in Simultaneous Bilingualism*. John Benjamins Publishing Company.

Hulk, A., & Müller, N. (2000). Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: Language and Cognition*, *3*(3), 227–224. https://doi.org/10.1017/S1366728900000353

Hulk, A., & van der Linden, E. (1996). Language Mixing in a French-Dutch Bilingual Child. In E. Kellerman, B. Weltens, & T. Bongaerts (Eds.), *Eurosla 6: a selection of papers* (pp. 89–101). Vereniging voor Toegepaste Taalwetenschap.

Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive Psychology*, *45*(3), 337–374. https://doi.org/10.1016/S0010-0285(02)00500-5

Jackendoff, R. S. (1972). *Semantic interpretation in generative grammar*. The MIT Press.

Jacobs, J. (2001). The dimensions of topic-comment. *Linguistics*, *39*(4), 641–681.

Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology*, *14*, 5–81.

Jia, G., Chen, J., Kim, H., Chan, P.-S., & Jeung, C. (2014). Bilingual lexical skills of school-age children with Chinese and Korean heritage languages in the United States. *International Journal of Behavioral Development*, *38*(4), 350–358. https://doi.org/10.1177/0165025414533224

Jia, G., & Fuse, A. (2007). Acquisition of English Grammatical Morphology by Native Mandarin-Speaking Children and Adolescents: Age-Related Differences. *Journal of Speech, Language, and*

Hearing Research, *50*, 1280–1299. https://doi.org/10.1044/1092-4388(2007/090

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60–99. https://doi.org/10.1016/0010-0285(89)90003-0

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 329–354.

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology*, *47*(2), 310–339. https://doi.org/10.1037/h0078820

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583–1585.

Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, *94*, 113–147. https://doi.org/10.1016/j.cognition.2004.01.002

Kaufman, A. S., Kaufman, N. L., & Melchers, P. (2003). *Kaufman Assessment Battery for Children: K-ABC*. PITS.

Kirk, N. W., Fiala, L., Scott-Brown, K. C., & Kempe, V. (2014). No evidence for reduced Simon cost in elderly bilinguals and bidialectals. *Journal of Cognitive Psychology*, *26*(6), 640–648. https://doi.org/10.1080/20445911.2014.929580

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events. *Cognition*, *95*, 95–127. https://doi.org/10.1016/j.cognition.2004.03.002

Kohn, M. L. (1963). Social Class and Parent-Child Relationships: An Interpretation. *American Journal of Sociology*, *68*(4), 471–480.

Kousaie, S., Sheppard, C., Lemieux, M., Monetta, L., Taler, V., Barker, L. A., Morgan, J., & Adam, R. (2014). Executive function and bilingualism in young and older adults. *Frontiers in Behavioral Neuroscience*, *8*(250), 1–12. https://doi.org/10.3389/fnbeh.2014.00250

Kovács, Á. M., & Mehler, J. (2009). Flexible learning of multiple speech structures in bilingual infants. *Science*, *325*(5940), 611–612. https://doi.org/10.1126/science.1173947

Kret, M. E., & Sjak-Shie, E. E. (2018). Preprocessing pupil size data: Guidelines and code. *Behavior Research Methods*, *51*, 1336–1342. https://doi.org/10.3758/s13428-018-1075-y

Kroll, J. F. (2008). Juggling Two Languages in One Mind. *Psychological Science Agenda*, *22*(1), 1–6. https://doi.org/10.1016/b978-0-12-394393-4.00007-8

Kroll, J. F., Dussias, P. E., Bogulski, C. A., & Kroff, J. R. V. (2012). Juggling two languages in one mind. What bilinguals tell us about language processing and its consequences for cognition. *Psychology of Learning and Motivation*, *56*, 229–262. https://doi.org/10.1016/B978-0-12-394393-4.00007-8

Kupisch, T. (2007). Determiners in bilingual German-Italian children: What they tell us about the relation between language influence and language dominance. *Bilingualism: Language and Cognition*, *10*(1), 57–78. https://doi.org/10.1017/S1366728906002823

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13.

Lago, S., Mosca, M., & Stutter Garcia, A. (2021). The Role of Crosslinguistic Influence in Multilingual Processing: Lexicon Versus Syntax. *Language Learning*, *71*, 163–192.

https://doi.org/10.1111/LANG.12412

Lemmerth, N., & Hopp, H. (2019). Gender processing in simultaneous and successive bilingual children: Cross-linguistic lexical and syntactic influences . *Language Acquisition*, *26*(1), 21–45. https://doi.org/10.1080/10489223.2017.1391815

Lenhard, A., Lenhard, W., Segerer, R., & Suggate, S. P. (2015). *PeabodyPicture Vocabulary Test - Revision IV (German Adaption)*. Pearson Assessment.

Lew-Williams, C. (2017). Specific Referential Contexts Shape Efficiency in Second Language Processing: Three Eye-Tracking Experiments with 6- and 10-Year-Old Children in Spanish Immersion Schools. *Annual Review of Applied Linguistics*, *37*, 128–147. https://doi.org/10.1017/S0267190517000101

Lum, J. A. G., Youssef, G. J., & Clarka, G. M. (2017). Using pupillometry to investigate sentence comprehension in children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, *60*(6), 1648–1660. https://doi.org/10.1044/2017_JSLHR-L-16-0158

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Syntactic Ambiguity Resolution as Lexical Ambiguity Resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 123–154). Lawrence Erlbaum.

Mack, J. E., Meltzer-Asscher, A., Barbieri, E., & Thompson, C. K. (2013). Neural Correlates of Processing Passive Sentences. *Brain Sciences*, *3*, 1198–1214. https://doi.org/10.3390/brainsci3031198

Macnamara, J. (1966). *Bilingualism and primary education*. Edinburgh University Press.

MacWhinney, B. (2007). A Unified Model. In N. Ellis & P. Robinson (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 1–43). Lawrence Erlbaum.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior.*, *23*(2), 127–150. https://doi.org/10.1016/S0022-5371(84)90093-8

Manetti, C. (2013). On the production of passives in Italian: evidence from an elicited production task and a syntactic priming study with preschool children. In Sarah Baiz, N. Goldman, & R. Hawkes (Eds.), *Proceedings of the 37th annual Boston University Conference on Language Development [BUCLD37]* (p. 2012). Cascadilla Press. http://www.bu.edu/bucld/supplementvol37/

Maratsos, M., Fox, D. E. C., Becker, J. A., & Chalkley, M. A. (1985). Semantic restrictions on children's passives. *Cognition*, *19*, 167–191. https://doi.org/10.1016/0010-0277(85)90017-4

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language and Hearing Research*, *50*(4), 940–967.

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, *6*(2), 97–115. https://doi.org/10.1017/s1366728903001068

Marinis, T. (2007). On-line Processing of Passives in L1 and L2 Children. In A. Belikova (Ed.), *Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition North America (GALANA)* (pp. 265–276). Cascadilla Press.

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *11*(1), 81–93. https://doi.org/10.1017/S1366728907003227

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*, 314–324. https://doi.org/10.3758/s13428-011-0168-7

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*(4), 372–380. https://doi.org/10.3758/BF03206780

Mazuka, R., Jincho, N., & Oishi, H. (2009). Development of executive control and language processing. *Linguistics and Language Compass*, *3*(1), 59–89. https://doi.org/10.1111/j.1749-818X.2008.00102.x

McDonald, J. L. (1986). The development of sentence comprehension strategies in English and Dutch. *Journal of Experimental Child Psychology*, *41*, 317–335. https://doi.org/10.1016/0022-0965(86)90043-3

Meir, N., & Armon-lotem, S. (2017). Independent and Combined Effects of Socioeconomic Status (SES) and Bilingualism on Children's Vocabulary and Verbal Short-Term Memory. *Frontiers in Psychology*, *8*(1442), 1–12. https://doi.org/10.3389/fpsyg.2017.01442

Meisel, J. M. (1986). Word order and case marking in early child language. Evidence from simultaneous acquisition of two first languages: French and German. *Linguistics*, *24*, 123–183.

Meisel, J. M. (1989). Early differentiation of languages in bilingual children. In K. Hyltenstam & L. Obler (Eds.), *Bilingualism Across the Lifespan: Aspects of Acquisition, Maturity and Loss*. Cambridge University Press.

Meisel, J. M. (2011). *First and Second Language Acquisition: Parallels and Differences*. Cambridge University Press. https://doi.org/10.1017/ CBO9780511862694

Meisel, Jürgen M. (2009). Second language acquisition in early childhood. *Zeitschrift Fur Sprachwissenschaft*, *28*, 5–34. https://doi.org/10.1515/ZFSW.2009.002

Melchers, P., & Melchers, M. (2015). *KABC-II. Kaufman Assessment Battery for Children – II*. Pearson Assessment.

Meyer, A. M., Mack, J. E., & Thompson, C. K. (2012). Tracking Passive Sentence Comprehension in Agrammatic Aphasia. *Journal of Neurolinguistics*, *25*(1), 31–43. https://doi.org/10.1016/j.jneuroling.2011.08.001.Tracking

Minai, U., Jincho, N., Yamane, N., & Mazuka, R. (2012). What hinders child semantic computation: Children's universal quantification and the development of cognitive control. *Journal of Child Language*, *39*(5), 919–956. https://doi.org/10.1017/S0305000911000316

Mishra, R. K. (2018). *Bilingualism and Cognitive Control*. Springer International Publishing AG. https://doi.org/10.1007/978-3-319-92513-4

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science*, *10*(6), 719–726. https://doi.org/10.1111/j.1467-7687.2007.00623.x

Müller, G. (1999). Optimality, markedness, and word order in German. *Linguistics*, *37*, 777–818. https://doi.org/10.1515/ling.37.5.777

Müller, N. (1998). Transfer in bilingual first language acquisition. *Bilingualism: Language and Cognition*, *1*(3), 151–171.

Müller, N. (2016). *Mehrsprachigkeitsforschung*. Narr Francke Attempo Verlag GmbH + Co. KG.

Müller, N. (2017). Different sources of delay and acceleration in early child bilingualism. *Zeitschrift Für Sprachwissenschaft*, *36*(1), 7–30. https://doi.org/10.1515/zfs-2017-0002

Müller, N., Cantone, K., Kupisch, T., & Schmitz, K. (2002). Zum Spracheneinfluss im bilingualen Erstspracherwerb: Italienisch – Deutsch. *Linguistische Berchte*, *190*, 157–206.

Müller, N., & Hulk, A. (2001). Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism: Language and Cognition*, *4*(1), 1–21.

Müller, N., & Hulk, A. (2000). Crosslinguistic influence in bilingual children: object omissions and Root Infinitives. In C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University Conference on Child Language Development* (pp. 546–557). Cascadilla Press.

Müller, N., Kupisch, T., Schmitz, K., & Cantone, K. (2007). *Einführung in die Mehrsprachigkeitsforschung* (2. Auflage). Narr Francke Attempo Verlag GmbH + Co. KG.

Nicoladis, Elena, & Genesee, F. (1997). Language development in preschool bilingual children. *Journal of Speech-Language Pathology and Audiology*, *21*, 258–270.

Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2013). 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., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective and Behavioral Neuroscience*, *5*(3), 263–281. https://doi.org/10.3758/CABN.5.3.263

Ong, C., Hutch, M., & Smirnakis, S. (2019). The Effect of Ambient Light Conditions on Quantitative Pupillometry. *Neurocritical Care*, *30*(2), 316–321. https://doi.org/10.1007/s12028-018-0607-8

Ostadghafour, S., & Bialystok, E. (2021). Comprehension of complex sentences with misleading cues in monolingual and bilingual children. *Applied Psycholinguistics*, *42*, 1117–1134. https://doi.org/10.1017/S0142716421000138

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*, 232–258. https://doi.org/10.1016/j.cogpsych.2012.12.002

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, *69*, 265–278. https://doi.org/10.1016/j.cortex.2015.04.014

Paap, K. R., Johnson, H. A., & Sawi, O. (2016). Should the search for bilingual advantages in executive functioning continue? *Cortex*, *74*, 305–314. https://doi.org/10.1016/j.cortex.2015.09.010

Paradis, J. (2011). Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism*, *3*(1), 213–237. https://doi.org/10.1075/lab.1.3.01par

Paradis, J., & Genesee, F. (1996). Syntactic Acquisition in bilingual children: Autonomous or Interdependent? *Studies in Second Language Acquisition*, *18*, 1–15. https://doi.org/10.1017/S0272263100014662

Peal, E., & Lambert, W. (1962). The relation of bilingualism to intelligence. *Psychological Monographs*, *76*(546), 1–23.

Petersen, J. (1988). Word-internal code-switching constraints in a bilingual child's grammar. *Linguistics*, *26*, 479–493.

Place, S., & Hoff, E. (2011). Properties of Dual Language Exposure That Influence 2-Year-Olds' Bilingual Proficiency. *Child Development*, *82*(6), 1834–1849. https://doi.org/10.1111/j.1467-

8624.2011.01660.x

Platzack, C. (1999). The vulnerable C-domain. *Paper Presented at the Workshop on Language Acquisition and Language Breakdown, Utrecht.*

Platzack, C. (2001). The vulnerable C-domain. *Brain and Language*, *77*, 364–377. https://doi.org/10.1006/brln.2000.2408

Pontikas, G. (2019). *Language Processing in Bilingual Children and Adults: Evidence from Filler-Gap Dependencies and Garden Path Sentences.* [Doctoral dissertation. University of Reading]. https://core.ac.uk/download/pdf/286352835.pdf

Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, *13*(2), 253–262. https://doi.org/10.1017/S1366728909990526

R Core Team. (2022). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. https://www.r-project.org/

Raven, J. C. (2003). *Raven's Coloured Progressive Matrices (CPM).* Pearson Publishing.

Richtsmeier, P. T., Gerken, L., Goffman, L., & Hogan, T. (2009). Statistical frequency in perception affects children's lexical production. *Cognition*, *111*, 372–377. https://doi.org/10.1016/j.cognition.2009.02.009

Rizzi, L. (1990). *Relativized Minimality*. The MIT Press.

Rizzi, L. (1997). The Fine Structure of the Left Periphery. In H. L. (Ed.), *Elements of Grammar*. Kluwer International Handbooks of Linguistics. Springer. https://doi.org/10.1007/978-94-011-5420-8_7

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

Rothweiler, M. (2006). The acquisition of V2 and subordinate clauses in early successive acquisition of German. In C. Lleó (Ed.), *Interfaces in Multilingualism: Accquisition and Representation* (pp. 91–113). John Benjamins Publishing Company. https://doi.org/10.1075/hsm.4.05rot

RStudio Team. (2022). *RStudio: Integrated Development for R.* RStudio, Inc. http://www.rstudio.com/

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia*, *42*, 1029–1040. https://doi.org/10.1016/j.neuropsychologia.2003.12.012

Russi, C. (2008). *Italian Clitics - An empirical Study*. Walter de Gruyter.

Saer, D. J. (1923). The Effect of Bilingualism on Intelligence. *British Journal of Psychology*, *14*(1), 25–38.

Sauermann, A. (2016). *Impact of the type of referring expression on the acquisition of word order variation.* [Doctoral dissertation. University of Potsdam]. https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/deliver/index/docId/8940/file/sps09.pdf

Sauermann, A., & Gagarina, N. (2017). The roles of givenness and type of referring expression in the comprehension of word order in Russian-speaking children. *Linguistics Vanguard*, *20160094*, 1–10. https://doi.org/10.1515/lingvan-2016-0094

Scheepers, C., Hemforth, B., & Konieczny, L. (2000). Linking syntactic functions with thematic roles: Psych-verbs and the resolution of subject-object ambiguity. In B. Hemforth & L. Konieczny (Eds.), *German sentence processing* (pp. 95–136). Kluwer Academic Publishers.

Schipke, C. S., Knoll, L. J., Friederici, A. D., & Oberecker, R. (2012). Preschool children's interpretation of object-initial sentences: Neural correlates of their behavioral performance. *Developmental Science*, *15*(6), 762–774. https://doi.org/10.1111/j.1467-7687.2012.01167.x

Schmitz, K. (2007). L'interface Syntaxe-Pragmatique: Le sujet chez des enfants bilingues francoallemands et italo-allemands. *AILE* , *25*, 9–43.

Schulz, P., & Grimm, A. (2019). The Age Factor Revisited: Timing in Acquisition Interacts With Age of Onset in Bilingual Acquisition. *Frontiers in Psychology*, *9*(2732), 1–18. https://doi.org/10.3389/fpsyg.2018.02732

Schumacher, P. B., & Hung, Y.-C. (2012). Positional influences on information packaging: Insights from topological fields in German. *Journal of Memory and Language*, *67*, 295–310. https://doi.org/10.1016/j.jml.2012.05.006

Schwartz, B. D. (2004). Why Child L2 acquisition? In J. Van Kampen & S. Baauw (Eds.), *Proceedings of Generative Approaches to Language 2003* (pp. 47–66). LOT Occasional Series. https://doi.org/10.1075/hsm.1.10kat

Sekerina, I., Stromswold, K., & Hestvik, A. (2004). How do adults and children process referentially ambiguous pronouns? *Journal of Child Language*, *31*, 123–152. https://doi.org/10.1017/s0305000903005890

Serratrice, L. (2013). Cross-linguistic influence in bilingual development: Determinants and mechanisms. *Linguistic Approaches to Bilingualism*, *3*(1), 3–25. https://doi.org/10.1075/lab.3.1.01ser

Serratrice, L., & Sorace, A. (2002). Overt and null subjects in monolingual and bilingual Italian acquisition. In B. Beachley, A. Conlin, & F. Brown (Eds.), *Proceedings of the 27th Boston University Conference on Child Language Development.* (pp. 739–750). Cascadilla Press.

Serratrice, L., Sorace, A., & Paoli, S. (2004). Crosslinguistic influence at the syntax-pragmatics interface: Subjects and objects in English-Italian bilingual and monolingual acquisition. *Bilingualism: Language and Cognition*, *7*(3), 183–205. https://doi.org/10.1017/S1366728904001610

Snedeker, J. (2013). Children's sentence processing. In R. P. G. Van Gompel (Ed.), *Sentence Processing* (pp. 189–220). Taylor & Francis.

Snedeker, J., & Geren, J. (2007). Starting Over: International Adoption as a Natural Experiment in Language Acquisition. *Psychological Science* , *18*, 79–87.

Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguistic Approaches to Bilingualism*, *1*(1), 1–33. https://doi.org/10.1075/lab.1.1.01sor

Souza, A. L., Byers-Heinlein, K., Poulin-Dubois, D., Hannon, E. E., Floccia, C., & Van Heugten, M. (2013). Bilingual and monolingual children prefer native-accented speakers. *Frontiers in Psychology*, *4*(953), 1–6. https://doi.org/10.3389/fpsyg.2013.00953

Speyer, A. (2007). Die Bedeutung der Centering Theory für Fragen der Vorfeldbesetzung im Deutschen. *Zeitschrift Für Sprachwissenschaft*, *26*, 83–115. https://doi.org/10.1515/ZFS.2007.011

SR Research Data Viewer 4.1.1. (2019). *[Computer software]*. SR Research Ltd.

SR Research Experiment Builder 2.1.140. (2017). *[Computer software]*. SR Research Ltd.

Stella, G., Pizzoli, C., & Tressoldi, P. E. (2000). *Il Peabody Test. Test di vocabolario recettivo.* Omega Edizioni.

Struys, E., Duyck, W., & Woumans, E. (2018). The Role of Cognitive Development and Strategic Task Tendencies in the Bilingual Advantage Controversy. *Frontiers in Psychology*, *9*(1970), 1–11. https://doi.org/10.3389/fpsyg.2018.01790

Struys, E., Mohades, G., Bosch, P., & Van Den Noort, M. (2015). Cognitive Control in Bilingual

Children: Disentangling the Effects of Second-Language Proficiency and Onset Age of Acquisition. *Swiss Journal of Psychology*, *74*(2), 65–73. https://doi.org/10.1024/1421-0185/a000152

Suraniti, S., Ferri, R., & Neri, V. (2009). *Test for Reception of Grammar, version 2 (TROG-2)*. Giunti OS.

Teubner-Rhodes, S., Mishler, A., Corbett, R., Andreu, L., Sanz-Torrent, M., Trueswell, J. C., & Novick, J. M. (2016). The effects of bilingualism on conflict monitoring, cognitive control, and garden-path recovery. *Cognition*, *150*, 213–231. https://doi.org/10.1016/j.cognition.2016.02.011

Thordardottir, E. (2014). The typical development of simultaneous bilinguals: Vocabulary, morphosyntax and language processing in two age groups of Montreal preschoolers. In T. Grüter & J. Paradis (Eds.), *Input and Experience in Bilingual Development* (pp. 141–130). John Benjamins Publishing Company.

Thordardottir, E., Rothenberg, A., Rivard, M.-E., & Naves, R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages? *Journal of Multilingual Communication Disorders*, *4*(1), 1–21. https://doi.org/10.1080/14769670500215647

Torregrossa, J., Andreou, M., Bongartz, C., & Tsimpli, I. M. (2021). Bilingual acquisition of reference: The role of language experience, executive functions and cross-linguistic effects. *Bilingualism: Language and Cognition*, *24*, 694–706. https://doi.org/10.1017/S1366728920000826

Traxler, M. J., Corina, D. P., Morford, J. P., Hafer, S., & Hoversten, L. J. (2014). Deaf readers response to syntactic complexity: Evidence from self-paced reading. *Memory and Cognition*, *42*, 97–111.

Trueswell, J. C., & Gleitman, L. R. (2004). Children’s eye-movements during listening: Evidence for a constraint-based theory of parsing and word learning. In J. M. Henderson & F. Ferreira (Eds.), *Interface of vision, language, and action* (pp. 319–346). Psychology Press.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*(2), 89–134. https://doi.org/10.1016/S0010-0277(99)00032-3

Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a lexicalist framework of constraint-based syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 155–180). Lawrence Erlbaum.

Tsimpli, I. M. (2014). Early, late or very late? Timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism*, *4*(3), 283–313. https://doi.org/10.1075/lab.4.3.01tsi

Tsimpli, I. M., Vogelzang, M., Balasubramanian, A., Marinis, T., Alladi, S., Reddy, A., & Panda, M. (2020). Linguistic diversity, multilingualism, and cognitive skills: A study of disadvantaged children in India. *Languages*, *5*(10), 1–22. https://doi.org/10.3390/languages5010010

Unsworth, S. (2013). Assessing the role of current and CUMULATIVE exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition*, *16*(1), 86–110. https://doi.org/10.1017/S1366728912000284

Unsworth, S. (2015). Amount of exposure as a proxy for dominance in bilingual language acquisition. In C. Silva-Corvalan & J. Treffers-Daller (Eds.), *Language dominance in bilinguals: Issues of measurement and operationalization* (pp. 156–173). Cambridge University Press. https://doi.org/10.1017/CBO9781107375345.008

Unsworth, S. (2016a). Quantity and quality of language input in bilingual language development. In E. Nicoladis & S. Montanari (Eds.), *Bilingualism across the lifespan: Factors moderating language proficiency* (pp. 103–121). American Psychological Association. https://doi.org/10.1037/14939-007

Unsworth, S. (2016b). Early child L2 acquisition: Age or input effects? Neither, or both? *Journal of*

Child Language, *43*(3), 608–634. https://doi.org/10.1017/S030500091500080X

Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A., & Tsimpli, I. M. (2014). The role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics*, *35*, 765–805. https://doi.org/10.1017/S0142716412000574

Unsworth, S., Brouwer, S., Bree, E. De, & Verhagen, J. (2019). Predicting bilingual preschoolers' patterns of language development : Degree of non-native input matters. *Applied Psycholinguistics*, *40*, 1189–1219. https://doi.org/10.1017/S0142716419000225

Unsworth, S., Hulk, A., & Marinis, T. (2011). Internal and external factors in child second language acquisition. *Linguistic Approaches to Bilingualism*, *1*(3), 207–212. https://doi.org/10.1075/lab.1.3.00int

Van Den Noort, M., Struys, E., Bosch, P., Jaswetz, L., Perriard, B., Yeo, S., Barisch, P., Vermeire, K., Lee, S.-H., & Lim, S. (2019). Does the Bilingual Advantage in Cognitive Control Exist and If So, What Are Its Modulating Factors? A Systematic Review. *Behavioral Sciences (Basel)*, *9*(27), 1–30. https://doi.org/10.3390/bs9030027

van der Linden, E., & Sleeman, P. (2007). Clitic Dislocation: Evidence for a low Topic position. *Linguistics in the Netherlands*, 173–186.

Van Engen, K. J., & McLaughlin, D. J. (2018). Eyes and ears: Using eye tracking and pupillometry to understand challenges to speech recognition. *Hearing Research*, *369*, 56–66. https://doi.org/10.1016/j.heares.2018.04.013.Eyes

Van Gompel, R. P. G., & Arai, M. (2018). Structural priming in bilinguals . *Bilingualism: Language and Cognition*, *21*(3), 448–455. https://doi.org/10.1017/S1366728917000542

van Rij, J., Hendriks, P., van Rijn, H., Baayen, H., & Wood, S. N. (2019). Analyzing the Time Course of Pupillometric Data. *Trends in Hearing*, *23*, 1–22. https://doi.org/10.1177/2331216519832483

Vasić, N., Chondrogianni, V., Marinis, T., & Blom, E. (2012). Processing of gender in Turkish-Dutch and Turkish-Greek child L2 learners. In A. K. Biller, E. Y. Chung, & A. E. Kimball (Eds.), *BUCLD36: proceedings of the 36th annual Boston University Conference on Language Development* (pp. 646–659). Cascadilla Press. https://dspace.library.uu.nl/bitstream/handle/1874/294408/BUCLDproceedings_2011_final_version.pdf?sequence=1

Vasilyeva, M., Waterfall, H., Gámez, P. B., Gómez, L. E., Bowers, E., & Shimpi, P. (2010). Cross-linguistic syntactic priming in bilingual children. *Journal of Child Language*, *37*, 1047–1064. https://doi.org/10.1017/S0305000909990213

Vikner, S. (1994). Finite verb movement in Scandinavian embedded clauses. In D. Lightfoot & N. Hornstein (Eds.), *Verb movement* (pp. 117–148). Cambridge University Press.

Volpato, F., Verin, L., & Cardinaletti, A. (2016). The comprehension and production of verbal passives by Italian preschool-age children. *Applied Psycholinguistics*, *37*, 901–931. https://doi.org/10.1017/S0142716415000302

Volterra, V., & Taeschner, T. (1978). The acquisition and development of language by bilingual children. *Journal of Child Language*, *5*(2), 311–326. https://doi.org/10.1017/S0305000900007492

Vuong, L. C., & Martin, R. C. (2015). The role of LIFG-based executive control in sentence comprehension. *Cognitive Neuropsychology*, *32*(5), 243–265. https://doi.org/10.1080/02643294.2015.1057558

Wendt, D., Brand, T., & Kollmeier, B. (2014). An eye-tracking paradigm for analyzing the processing time of sentences with different linguistic complexities. *PLoS ONE*, *9*(6), 1–13.

https://doi.org/10.1371/journal.pone.0100186

Wiseheart, M., Viswanathan, M., & Bialystok, E. (2016). Flexibility in task switching by monolinguals and bilinguals . *Bilingualism: Language and Cognition*, *19*(1), 141–146. https://doi.org/10.1017/S1366728914000273

Wittek, A., & Tomasello, M. (2005). German-speaking children's productivity with syntactic constructions and case morphology: Local cues act locally. *First Language*, *25*(1), 103–125. https://doi.org/10.1177/0142723705049120

Yang, H., Hartanto, A., & Yang, S. (2018). Bilingualism confers advantages in task switching: Evidence from the dimensional change card sort task . *Bilingualism: Language and Cognition*, *21*(5), 1091–1109. https://doi.org/10.1017/S136672891700044X

Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The Pupil Dilation Response to Auditory Stimuli: Current State of Knowledge. *Trends in Hearing*, *22*, 1–25. https://doi.org/10.1177/2331216518777174

Zelazo, P. D., & Frye, D. (1997). Cognitive complexity and control: A theory of the development of deliberate reasoning and intentional action. In M. Stamenov (Ed.), *Language structure, discourse, and the access to consciousness* (pp. 113–153). John Benjamins Publishing Company.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*, 37–63. https://doi.org/10.1016/S0885-2014(96)90027-1

Declaration of authorship

I hereby certify that the thesis I am submitting is entirely my own original work except where otherwise indicated. I am aware of the University’s regulations concerning plagiarism, including those regulations concerning disciplinary actions that may result from plagiarism. Any use of the works by any other author, in any form, is properly acknowledged at their point of use.

Potsdam, January 2023

Chiara Boila