The cognitive and linguistic abilities of bilinguals with genetic disorders: the Prader-Willi Syndrome population

Estela Garcia Alcaraz

Thesis submitted to the University of Ottawa
in partial fulfillment of the requirements for the
Doctorate in Philosophy degree in Spanish

Department of Modern Languages and Literatures
Faculty of Arts
University of Ottawa

Acknowledgements

Little could I imagine a few years ago how enriching the process of carrying out this dissertation would be, both personally and professionally, from its inception to its defence. I will always be grateful to Dr. Liceras for suggesting that I expand my field of expertise by investigating the effects of bilingualism in populations with atypical language development due to genetic syndromes. From the very first moment I was captivated by the idea and we started to work together in this understudied, although growing, field of research, which has become my main research interest.

As much as I can try, I will never be able to express with words what working with Dr. Liceras has meant to me both professionally and personally. It is no secret that Dr. Liceras is a renowned scholar within the field of bilingualism and language acquisition, but what some may not know is that she is not only an extraordinary researcher but also an extraordinary and very generous supervisor. However, no matter how good these qualities are, her quality that I value the most, without any doubt, is her humanity, since she is one of the best people I have ever had the chance to meet. Juana, thank you for agreeing to work with me for all these years and for giving me the chance to feel at home, although being far away from it, with you and your family. I will always treasure the moments shared with Vicente, since he inspired me more than he could imagine and for teaching me what it means to be principled in life. I know that he would be very happy to see this project completed, which would have not been possible without Ivo, its driving force and the reason behind it. Thank you, Ivo, for being always willing to help, and for always having a smile on your face ready to get involved. It has been an absolute privilege to count on you. Also, thank you Marco for your always incisive comments and your careful reading of this dissertation.

I am very grateful to the members of my thesis examining committee for all their insightful feedback and suggestions: Dr. Kay Raining-Bird, Dr. Slavkov, Dr. Sutton, and Dr. Valenzuela. Dr. Slavkov and Dr. Sutton, thank you very much for accompanying me in my journey from the very beginning, and for all your help, dedication and encouragement all these years. Your commitment has been beyond expectations. I would also like to thank Dr. Desrochers for discovering all the E-prime potential to me and for his generosity discussing my online experiments. Thank you also to Dr. Vinerte for being always willing to share your time and expertise with me on the executive control task design. Dr. Alves-Soares deserves special mention since he was the one who introduced me to the fascinating world of mixed-model statistical analysis; together we have experienced very frustrating but also very rewarding moments with our respective statistical analyses. Likewise, thank you to the different members of the *Language Acquisition Research Laboratory* of the University of Ottawa. It has been a pleasure to share these years with you in such a stimulating group of junior and senior researchers.

To carry out this dissertation would have not been possible without the funding received from the Ontario Graduate Scholarship program (Government of Ontario and the University of Ottawa) and from the Doctoral Research Scholarship program (Fonds de Recherche du Québec-Société et Culture). Also, different travel and research grants from the APTPUO, CUPE 2626 and the Faculty of Arts of the University of Ottawa have allowed me to share part of my dissertation in different international conferences and to collect data in Spain. Administratively, I would like to thank Philippe Villeneuve for being so diligent and for making all the arduous administrative processes so easy.

The real protagonists of this dissertation –all the participants and their parents– deserve a special mention. It has been an absolute honour to know you all and to share time with you. Data collection has been by far one of the most rewarding experiences of all this process due to your willingness to be part of this project and your contagious enthusiasm. Data collection would have not been possible without the help of Assumpta Mas, Aurora Rustarazo, and Carme Brun, as well as both the *Asociación Española Síndrome Prader Willi* and the *Asociación Prader-Willi Cataluña.* For this and more, thank you very much!

On a more personal note, I would like to thank my family and friends, especially Ester Ballester and Irene Guzmán for agreeing on sharing their time and knowledge as narrative evaluation referees, and to my Canadian family for being always there for me: David García León, Javier García León, Nelson Méndez, and Raquel Llama.

Last but not least, my warmest gratitude to the two most important people in my life: my mother and Fabio. *Gracias mamá por apoyarme siempre en lo que quiero hacer sin peros ni condiciones y por caminar siempre a mi lado. Nunca me he sentido sola ni lejos de ti aun estando en continentes distintos.* Fabio, you have been the most beautiful thing that these years have brought me. You have always been my light and my comfort and I could have never asked for a better partner to walk with through this journey. Now, a new life is ahead of us and I am looking forward to writing the next chapter together.

Abstract

The main objective of this dissertation is to investigate the cognitive and linguistic abilities of monolinguals and bilinguals with Prader-Willi Syndrome, a genetic disorder that, among other characteristics, generally involves mild to moderate intellectual disability (see Cassidy et al., 2012). More specifically, we investigate whether and how the fact of speaking two languages may shape the executive control, metalinguistic and narrative abilities of individuals with this syndrome.

The study of the effects of bilingualism among non-typically developing individuals is scarce and in its initial stages –being especially limited for individuals with intellectual disabilities (see Kay-Raining Bird et al., 2016 for an overview). Traditionally, non-typically developing individuals have been discouraged from becoming bilingual under the belief that speaking two languages could be counter-productive (Peña, 2016; among others). However, the limited research available, albeit growing, has revealed that these individuals not only become bilingual but also that speaking two languages does not seem to affect them negatively (Cleave et al., 2014; among others). These findings, together with the fact that previous research has mostly highlighted positive outcomes of bilingualism at the executive control, metalinguistic and narrative levels for the typically developing population (Andreou, 2015; Bialystok, 2001; Valian, 2015; among others), led us to explore whether monolingual and bilingual individuals with Prader-Willi Syndrome differed in the three aforementioned areas, and, if so, whether a potential bilingual advantage was attested among this population. We recruited seven Spanish-Catalan early sequential bilingual speakers (*M* = 18.04 years old, *SD* = 8.95, age range 10;5-33;10) and eight Spanish monolingual speakers (*M* = 22.80 years old, *SD* = 11.58, age range 9;4-47;0) with Prader-Willi Syndrome. Catalan was the dominant language for all bilingual speakers with the exception

of one participant, who declared Spanish as his/her dominant language, and both monolingual and bilingual groups were comparable in terms of nonverbal Intelligence Quotient, receptive vocabulary, and sentence recall abilities in Spanish. Participants were administered two executive control tasks, one language driven task (the Stroop task) and one non-language based task (the Flanker task); two metalinguistic tasks (a Sentence Judgment task and a Word-Length Judgment task), and an oral spontaneous narrative task elicited using the Frog Story series. Narratives were analyzed at both the macrostructural and microstructural levels and bilinguals were asked to complete the metalinguistic and the narrative tasks in both Spanish and Catalan. Results yielded similar outcomes for monolinguals and bilinguals at the executive control, metalinguistic and narrative macrostructural levels. However, bilinguals showed a relative advantage at the microstructural narrative level (more vocabulary variability). Thus, our findings not only suggest that individuals with Prader-Willi Syndrome can become bilingual without evidence of negative effects, which is in line with previous research, but also that they can achieve a similar level of performance, or even outperform monolingual speakers in certain linguistic abilities in Spanish – the non-dominant language for the majority of the bilingual participants. Spanish-Catalan bilinguals showed comparable metalinguistic and narrative abilities in both their languages.

Despite the fact that inferential statistical analyses did not reveal a significant and consistent bilingual advantage, descriptive data seems to point to it. With all due caution, these results suggest that the effects of bilingualism for the Prader-Willi Syndrome population may be somehow similar to those traditionally proposed for the typically developing population since, when an effect of bilingualism was found, this was positive. These findings are important and constitute a relevant contribution to the field, even though we acknowledge that they are preliminary and should be confirmed with further data.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AAIDD	American Association on Intellectual and Developmental Disabilities
AC	Analogical Criterion
AIC	Akaike Information Criterion
AICc	Second-order Akaike Information Criterion
ANOVA	Analysis of Variance
ANT	Attentional Network Test
AoO	Age of Onset
APA	American Psychiatric Association
ASD	Autism Spectrum Disorder
BA	Bilingual Advantage
BADS	Behavioral Assessment of Dysexecutive Syndrome
BAH	Bilingual Advantage Hypothesis
BEPA	Bilingual Executive Processing Advantage
BICA	Bilingual Inhibitory Control Advantage
BSL	British Sign Language
CELF	Clinical Evaluation of Language Fundamentals
CHAT	Codes for the Human Analysis of Transcripts
CHILDES	Child Language Data Exchange System
DCCS	Computerized Dimensional Change Card Sort task
DD	Developmental Disability
DEL	Deletion
DLD	Developmental Language Disorders
DS	Down Syndrome
EC	Executive Control
ERP	Event-Related Potential
GCSC	Grammatically Correct and Semantically Correct
GCSV	Grammatically Correct with Semantic Violation
GJT	Grammatical Judgment Task
GVSC	Grammar Violation but Semantically Correct

GVSV	Grammar Violation with Semantic Violation
ID	Intellectual Disability
IQ	Intelligence Quotient
L2	Second Language
MLU	Mean Length of Utterance
Non-TD	Non-Typically Developing
NSS	Narrative Scoring System
PPVT	Peabody Picture Vocabulary Test
PWS	Prader-Willi Syndrome
RT	Reaction Time
SCT	Sentence Completion Task
SES	Socioeconomic Status
SJT	Sentence Judgment Task
SLI	Specific Language Impairment
TD	Typically Developing
TONI	Test of Nonverbal Intelligence
TTR	Type-Token Ratio
UPD	Uniparental Disomy
WISC	Wechsler Intelligence Scale for Children
WLJT	Word-Length Judgment task

1. Introduction

It seems to be uncontroversial at this point in time that speaking more than one language is actually more common than speaking only one (Grosjean, 2010; Marian & Shook, 2012). Consequently, it is not surprising that bilingualism studies have become a fruitful and an autonomous research field, although, as Byers-Heinlein and Lew-Williams (2013) highlight, there are still many questions without answers that need to be addressed. However simple it may appear, one of the main challenges of the field has been to define bilingualism, a definition which is still a source of debate. Within the framework of this dissertation, and following Kay-Raining Bird et al.'s (2016, p. 2) approach, who in turn follow Grosjean's (1992) definition of bilingualism, we consider bilinguals those individuals who speak two languages and who alternate between them regularly.[1] This definition of bilingualism has been defended by Kay-Raining bird and colleagues as especially suitable when working with non-typically developing (henceforth non-TD) individuals, since the focus is put on the individuals' language use rather than on their language proficiency.

Bilingualism among the typically developing (henceforth TD) population has been widely scrutinized during the last decades and it has been suggested that speaking two languages comes with some "costs" and some "benefits". For instance, Ardila (2012) highlights that bilingual speakers, among other characteristics, may show a more reduced vocabulary in each one of their languages in comparison to monolingual speakers and that they may experience some language acquisition delays as well as crosslinguistic transfer. Conversely, Adesope et al. (2010) argue that bilingualism seems to enhance different cognitive and linguistic abilities, among them, but not

[1] Throughout this dissertation the terms bilingual speakers and dual speakers will be used interchangeably.

restricted to, executive control (henceforth EC) and metalinguistic abilities. These positive outcomes of bilingualism, specially at the EC level, have led some scholars to propose the existence of a bilingual advantage (henceforth BA). The bilingual advantage hypothesis (henceforth BAH)[2] (see Bialystok et al., 2012 for more details) has been extensively put to the test within the TD literature and, as we will discuss in the successive chapters, it has not been exempt from controversy. Some critical voices, especially during the last decade, have questioned the actual existence of such an advantage (Paap et al., 2015; Paap & Greenberg, 2013; among others). Despite the debate that may exist concerning the possible advantages or disadvantages that bilingualism may confer, there seems to be a general consensus on defending that, overall, bilingualism is more of an enriching than a harmful experience (Bialystok, 2016). Even if it were the case that bilingualism would not result in any cognitive or linguistic solid benefits, many bilingual speakers would agree that experiencing some small costs would be worth having access to two languages and two cultures as well as all the personal, affective, social, and professional implications that this may entail (see Bialystok, 2016 for a similar approach).

Taking this into account, it is important to highlight that while bilingualism is being defended and encouraged for TD individuals, some non-TD[3] individuals are still being discouraged, or at least not being encouraged enough, to become bilinguals (Kay-Raining Bird et al., 2012; Marinova-Todd et al., 2016; among others). Traditionally, individuals with developmental disabilities (henceforth DD), especially in those cases which also entail intellectual disabilities (henceforth ID), have been recommended to focus only on one language to avoid the

[2] The BAH is normally used in the literature to refer to an EC advantage for bilingual speakers. Nonetheless, in this dissertation both terms BA and BAH are applied likewise to the metalinguistic and narrative dimensions, since previous literature has also defended a recurrent positive effect of bilingualism at these two levels.
[3] Within the framework of this dissertation we will use the adjective non-typically developing to refer to both children and adults whose language development at present (the children) or during childhood (the adults) has not been "typical".

burden and the possible counter-productive effects on their first language development that adding a second language may entail (Cleave et al., 2014; Peña, 2016; among others). Nonetheless, since the beginning of this century, various researchers have devoted their research to investigating the effects of bilingualism in non-TD populations. This field of research has mainly focused on individuals with DD not entailing ID, as is the case of individuals with Developmental Language Disorders[4], or with the possibility to entailing ID, as it is the case of individuals with Autism Spectrum Disorder (henceforth ASD)[5], usually excluding those participants with an intelligence quotient (henceforth IQ) under 70, which is normally the cut-off point considered to indicate a potential ID. Thus, the study of bilingualism among individuals with ID remains practically unexplored and almost limited to Down syndrome (henceforth DS). Despite the idiosyncratic characteristics of these three populations, previous research has shown parallel outcomes: in the first place, individuals with DD become bilingual when given the opportunity and, in the second place, speaking two languages does not seem to affect them negatively (see Kay-Raining Bird et al., 2016 for an overview).

Educational professionals and practitioners appear to be aware of these research outcomes, but as Marinova-Todd et al. (2016) and Paradis (2016) argue, there seems to be a gap between theory (what the authors refer to as opinions) and practice. Nonetheless, we should recognize that substantial progress has been made during the last years to create a more inclusive school system accommodating students with different cognitive and linguistic profiles due to the needs of

[4] DLD was previously referred to as Specific Language Impairment (SLI) (see Bishop et al., 2017 for a terminological disscusion). From this moment onwards we will use the term DLD-SLI to refer to previous research focused on this DD. As defined in Bishop et al. (2017, p. 1078), DLD refers to "language problems that are severe enough to interfere with daily life, have a poor prognosis and are not associated with a clear biomedical aetiology".
[5] As defined by the American Psychiatric Association (2018), ASD "is a complex developmental condition that involves persistent challenges in social interaction, speech and nonverbal communication, and restricted/repetitive behaviors".

individuals with specific disabilities. Not long ago, individuals with DD and ID were usually excluded from mainstream schooling and had to attend special education programs where access to a second language (henceforth L2) was minimum, if not restricted. However, in most of the countries offering public and universal access to schooling, there is at present a general tendency to favour an inclusive education taking into account each individual's needs, which, in theory, increases the options to have access to an L2 (see Pesco et al., 2016 for an in depth analysis of inclusive education and L2 learning among individuals with DD). Nevertheless, and even though important and relevant progress has been made in the past years in this respect, more research is needed to make the field advance (Paradis, 2016), especially with regard to individuals with ID.

With this dissertation we want to contribute to this limited but growing body of research by investigating how monolinguals and bilinguals with genetic syndromes entailing ID compare at the cognitive and linguistic levels.[6] More specifically, we aim to shed light on how Spanish monolingual and early-sequential Spanish-Catalan bilingual speakers with Prader-Willi Syndrome (henceforth PWS) compare with respect to EC, metalinguistic and narrative abilities. Based on previous research investigating the effects of bilingualism in individuals with genetic syndromes (Cleave et al., 2014; Feltmate & Kay-Raining Bird, 2008; Kay-Raining Bird et al., 2005; among others), our working hypothesis is that PWS bilinguals would not show shortcomings when compared to their monolingual peers. In fact, we are also interested in elucidating whether dual speakers with PWS may show signs of benefitting from speaking two languages in a similar way to that proposed for the TD population. Putting the BAH to the test among non-TD individuals with DS (e.g., Martin, 2017), DLD-SLI (e.g., Peristeri et al., 2019) or ASD (e.g., Gonzalez-Barrero

[6] It is estimated that approximately between 1-3% of the population present ID and although different factors may cause it, genetic disorders have been proposed to be one of the main causes (Karam et al., 2015; Vissers et al., 2016).

& Nadig, 2019) is a new development in the field and we want to contribute to this line of research by providing new data from an understudied genetic syndrome. To this end, both monolingual and bilingual speakers were administered three standardized tests and four experiments combining both online (reaction time –RT) and offline (accuracy) methodologies as well as oral spontaneous data. While the main objective of the three standardized measures was to obtain information about the participants' cognitive and linguistic profile, the main goal of the different experimental tasks was to obtain information about whether and how both groups compared in three independent but interrelated areas where a BA has traditionally been suggested for the TD population.

During three different experimental sessions the participants were asked to complete a nonverbal IQ test, a receptive vocabulary test and a sentence recall test together with: (i) two EC tasks (experiment 1), a Flanker task and a color-word Stroop task; (ii) two metalinguistic tasks, a Sentence Judgment task (experiment 2) and a Word-length Judgment task (experiment 3); and (iii) a narrative task using the Frog Stories series as elicitation method (experiment 4).[7] Narratives were analyzed according to their macrostructure (coherence and organization) and their microstructure (specific linguistic form and content). Bilingual participants completed the standardized tests and the metalinguistic and narrative tasks in both their languages.

Besides the main objective previously presented, we set two additional objectives aiming at explaining whether individual differences in terms of nonverbal IQ, receptive vocabulary and sentence recall abilities may play a role on the overall participants' results, and whether bilingual speakers would display similar outcomes in their dominant language (Catalan for all but one of

[7] The different standardized tests and experimental tasks administered in this dissertation are presented and discussed in detail in chapters 4-7.

the bilinguals who recognized Spanish as his/her dominant language) and their non-dominant language.

This dissertation is the first of its kind and it contributes to the field at both the theoretical and experimental levels. At the theoretical level, it contributes novel data providing insights of how monolinguals and bilinguals with PWS compare at the EC, metalinguistic and narrative level and how bilingualism may interact with these abilities. This will allow researchers to have a more comprehensive view of how bilingualism may affect individuals with ID in different areas and whether the effects of speaking two languages are somehow comparable between the TD and the non-TD population. At the experimental level, it contributes data from an understudied language combination (Spanish-Catalan), an understudied type of bilingualism (early-sequential bilinguals), an understudied genetic disorder (PWS) and an understudied developmental stage (mainly adolescents and young adults). Likewise, it contributes data of different nature: experimental data using both online and offline methodologies and production data.

This dissertation is organized as follows:

Chapter 2 provides an overview of the effects of bilingualism at the EC, metalinguistic and narrative levels for both TD and non-TD individuals. The review for the non-TD population focuses on the two DD that have received more attention within the field: DLD-SLI and ASD. This chapter also contains a review of the main findings of previous research devoted to the study of bilingualism in individuals with genetic syndromes.

Chapter 3 includes a general snapshot of PWS with a special focus on these individuals' EC, metalinguistic and narrative abilities. This chapter also incorporates a review of two case studies conducted up to this day investigating the linguistic abilities of bilingual individuals with PWS.

In Chapter 4 we offer an overview of the study. We present the research questions and hypotheses as well as the participants, the methodology, the data collection procedure and the data analysis. The participants' section includes a detailed account of the participants' linguistic background as well as a comprehensive analysis of their cognitive and linguistic abilities. This information not only allows us to have a better picture of the participants' characteristics but also contributes to the field with new data, since studies centered on the cognitive and linguistic abilities of PWS individuals are relatively scant, especially involving Romance languages, and, to the best of our knowledge, they are nonexistent when it comes to bilingual speakers –at least in a group-study design. The results show that bilingual and monolingual speakers have comparable nonverbal IQ, receptive vocabulary and sentence recall abilities despite the fact that bilingual speakers are compared to monolinguals in their non-dominant language. A within-group analysis among bilingual speakers comparing their receptive vocabulary and their sentence recall abilities in both of their languages suggests better receptive vocabulary abilities in Catalan but comparable outcomes for sentence recall abilities in both Catalan and Spanish.

Chapter 5 reports the results (RT and accuracy data) of the EC experiment including the Flanker task and the Stroop task. Detailed information about the participants' sample, the experimental design, and the procedure, as well as the data analysis and the results is provided. The chapter concludes with a discussion of the results of both tasks, which show similar EC abilities for both monolingual and bilingual speakers with PWS.

Chapter 6 presents the results of the two metalinguistic tasks: the Sentence Judgment task, and the Word-length Judgment task. As in the previous chapter, and for each experiment, a description of the participants' sample, the experimental design, the procedure, the data analysis, and the results is provided. This chapter also includes the discussion of the results, which again do not reveal differences between monolingual and bilingual individuals with PWS. A within-group analysis for bilingual speakers shows that Spanish-Catalan speakers perform similarly in both tasks for both their languages.

Chapter 7 reports the results of the participants' narrative abilities. Monolinguals and bilinguals were asked to narrate *A boy, a dog, and a frog* (Mayer, 1967) in Spanish, and bilinguals were also asked to narrate *Frog, where are you?* (Mayer, 1969) in Catalan. Both narrative corpora were analyzed according to their macrostructure and microstructure. Monolinguals and bilinguals with PWS showed similar macrostructural abilities as well as comparable morphosyntactic language development; however, a BA was detected when evaluating participants' lexical variability. Again, no differences between languages were found for bilingual speakers.

Chapter 8 concludes this dissertation with a general discussion where the main findings for the cognitive and linguistic abilities of monolingual and bilingual individuals with PWS are summarized and the effects of bilingualism are examined. In addition, this chapter discusses the specific contributions, implications, and limitations of this dissertation, and provides recommendations for future research.

2. Bilingualism in typically and non-typically developing populations

The main aim of this chapter is to present an overview of the principal findings that previous research has revealed, for both TD and non-TD individuals, in relation to the interaction between bilingualism and EC, metalinguistic, and narrative abilities. It is beyond the scope of this dissertation to carry out an exhaustive review of each one of these three areas and we focus our review on a selection of representative studies that will allow us to provide an overall idea of the state of the field. The studies discussed here deal with more issues and sometimes have a broader scope than our depiction displays, however, we isolate what is relevant for our investigation; namely, we concentrate on the results obtained from similar tasks to the ones included in this dissertation, and on the population targeted in each case. In order to have a more comprehensive discussion for the TD population, the reader can consult Valian (2015) and Bialytsok (2018) for an overview of the effects of bilingualism on the EC domain, Bialystok (2001) and Roehr-Brackin (2018) for its effects on metalinguistics abilities, and Andreou (2015) for a review of the effects of bilingualism in narrative abilities. For a comprehensive review of bilingualism among non-TD individuals we direct the reader to Kay-Raining Bird et al. (2016) and to Paradis and Govindarajan (2018).

This chapter comprises three sections. In section 2.1 we present an outline of how bilingualism has been said to shape the EC, metalinguistic, and narrative abilities of the TD population, and, when possible, we review the results obtained from both children and adults. Studies centered on investigating the effects of bilingualism at the EC level have suggested that the developmental stage of the individual may condition the identification of a potential BA, since the effects of bilingualism seem to be broader in children than in the young adult population (see

Bialystok, 2018 for a review).[8] Thus, as will be further discussed in Chapter 3, since PWS is a genetic disorder entailing mild to moderate ID and there is typically a gap between these individuals' chronological age and their language development (Kleppe et al., 1990; Van Borsel et al., 2007), it will be interesting to determine which population our sample seems to be more in line with, as our participants mostly include adolescents and young adults. In section 2.2. we present a brief review of how bilingualism has been proposed to interact with the three cognitive and linguistic areas under study among DLD-SLI and ASD individuals. The fact, on the one hand, that DLD-SLI does not include individuals with ID and, on the other hand, that ASD studies within the field mainly focus on individuals without ID (Gonzalez-Barrero & Nadig, 2019; Meir & Novogrodsky, 2019; Peristeri et al., 2020; among others) justifies that we consider these individuals to belong to a different category and we will not study their profiles in detail. Nonetheless, previous findings related to both the TD and non-TD populations are relevant in order to establish the foundation of our study and to formulate our research questions and hypotheses in the light of bilingualism effects, which are detailed in chapter 4.

Lastly, in section 2.3, we include a succinct examination of the limited research available studying the effects of bilingualism in individuals with ID as a result of genetic syndromes. More specifically, in this section we revise previous findings for individuals with DS. PWS and DS are two genetic syndromes that despite the intrinsic characteristics that both conditions may have "share several clinical and functional features" (Cimolin et al., 2010, p. 2). Thus, with all due caution, our prediction is that individuals with DS and PWS would show, overall, more similarities than differences in relation to the effects caused by speaking two languages.

[8] The effects of bilingualism have also been defended to be larger in the older adult population (Bialystok, Poarch, et al., 2014; Gold et al., 2013).

2.1. Bilingualism in typically developing populations

2.1.1. Executive control abilities

One especially productive subfield within bilingualism studies has been the study of how bilingualism shapes EC abilities. EC, which has been also referred to as executive function or cognitive control (Diamond, 2013, p. 136), refers to those “general-purpose control mechanisms that enable people to regulate their thoughts and behaviors to align with their goals” (Friedman, 2016, p. 535). More specifically, as Sekerina et al. (2019, p. 2) point out, these mechanisms “refer to those cognitive processes that integrate, regulate, and control other cognitive processes, processes such as planning, inhibiting, shifting (from one task or rule to another), and updating (stored material with new material)”. It has been proposed that bilinguals would show certain advantages at the EC level in comparison to their monolingual peers due to their being accustomed to dealing with two languages and being thus somehow trained to ignore irrelevant information –in this case, the language not in use (Adesope et al., 2010; among others). In fact, inhibition has been one of the EC processes that has received more attention, and where the BA hypothesis has been widely put to the test. Nonetheless, it has been argued that it is important to distinguish between two different processes: “response inhibition” and “interference suppression” (Luk et al., 2010; Traverso et al., 2018; among others). In Brydges et al.’s (2013, p. 1) words, response inhibition denotes “the suppression of a prepotent or automatic behavioral response” and interference suppression indicates “the ability to control for distracting stimuli or information due to stimulus competition”. It is in the latter ability where previous research has found a potential BA.

Two tasks that have been extensively used to compare interference suppression abilities between monolingual and bilingual speakers are the Flanker task (Eriksen & Eriksen, 1974) and the Stroop task (Stroop, 1935). Very succinctly, the Flanker task is a non-language-based experiment in which participants are presented with different arrow/chevron sequences and are asked to determine the direction of the central arrow/chevron. A variation of this task is the Attentional Network Test (ANT–Fan et al., 2002), which combines a Flanker task and a Cue reaction time task (Posner, 1980). Conversely, in a language-based Stroop task participants are presented with different color words and are instructed to indicate the words' ink color.[9] Both tasks normally include congruent and incongruent trials and it is in the latter where participants are presented with conflict and are forced to suppress interference in order to succeed. Hilchey and Klein (2011) proposed two different hypotheses to capture the positive effects of bilingualism described in the literature, when detected: the "Bilingual Inhibitory Control Advantage" (BICA) and the "Bilingual Executive Processing Advantage" (BEPA) hypotheses. Under the BICA hypothesis, bilinguals are expected to outperform monolinguals when resolving conflict (incongruent condition) while under the BEPA hypothesis a general BA is expected across the board and not only on the incongruent condition. Although Hilchey and Klein's (2011) proposal was based on nonverbal EC task results, it seems reasonable to extrapolate their results to language driven tasks, as Coderre et al. (2014) or Vinerte (2018), among others, previously did.

We will now review how monolingual and bilingual children and adults compare when assessing their EC abilities using a Flanker task, an ANT task, and a Stroop-like task.

Kapa and Colombo (2013) administered the ANT task to two groups of Spanish-English bilinguals (L2 age of onset [AoO] before or after 3) and to an English monolingual group (age

[9] In this dissertation we focus on the word-color Stroop task design. However, there are other nonverbal proposals such as the picture-word Stroop task (MacLeod, 2006) or the numerical Stroop task (Hernández et al., 2010).

range 5;8-14;11 years old).[10] Overall, no differences between groups were detected but a general RT advantage was found for simultaneous bilinguals when the age and the verbal ability variables were controlled. According to the authors, these findings suggest that the AoO of bilingualism seems to play an important role in the EC abilities. Another factor that has been said to play an important role in EC is the participants' Socioeconomic Status (SES) (see Valian, 2015 for an overview). However, although SES may play a pivotal role in EC results, different studies have proved that when neutralizing this variable by including homogeneous participants, a BA can still arise. For example, Calvo & Bialystok (2014) administered the Flanker task to English monolingual and English-other language bilingual children who had relative proficiency in both of their languages and who had been exposed to English at least for two years (age range 6-7 years old). Participants belonged to either working or middle class families and results revealed that both factors, languages spoken and SES, played an important role in the EC abilities. On the one hand, accuracy results revealed a BA (no differences on RT data), and, on the other hand, they showed that children from middle class families were marginally more accurate than children from working class families. Similar results were found by Engel de Abreu et al. (2012) or Yang et al. (2011), among others, who found a BA when groups were matched on SES.

Nonetheless, despite the existence of different studies showing positive effects of bilingualism at the EC level, the literature is far from being straightforward. For example, Antón et al. (2014) administered the ANT task to Basque-Spanish bilingual (L2 AoO during early childhood) and Spanish monolingual children (mean age range 7.5-11.5) and no differences between groups were detected. Similar results were found by Bialystok et al. (2010), whose ANT

[10] Given that different researchers use several terms and cut-off points to refer to different types of bilingualism (e.g., early vs. late bilingualism), we have opted for specifying the bilinguals' L2 AoO, when it was provided. L2 within the framework of this dissertation refers to the non-dominant language or the language acquired later in time.

task results did not reveal a BA for English-other language bilinguals in comparison to their English monolingual counterparts (mean age range 3.0-4.5 years old).

Within the adult population the panorama is similar. Luk et al. (2011) compared what they call English-other language early bilinguals (L2 AoO before 10) and English-other language late bilinguals (L2 AoO after 10) to English monolinguals speakers. All participants were university students in their early twenties. They were administered a Flanker task and their results mirrored those of Kapa and Colombo's (2013) findings for children: better outcomes for early bilinguals when compared to late bilinguals and monolingual speakers and no differences between the last two groups. Pelham and Abrams (2014), on the other hand, found similar outcomes for both early and late English-Spanish bilinguals when compared to English monolinguals on the ANT task. As in the previous study, participants were university students in their early twenties but in this case the authors considered late bilinguals those individuals who became fluent in both of their languages after the age of 13 independently of bilingualism AoO. These findings, contrary to Luk et al. (2011) and Kapa and Colombo (2013), lead Pelham and Abrams (2014) to suggest that the effects of bilingualism would not be related to the age at which participants become bilinguals but to the frequency of use of both languages. Positive effects of bilingualism using the Flanker paradigm have also been found beyond behavioral data using the event-related functional magnetic resonance imaging technology (Abutalebi et al., 2012; Luk et al., 2010). As it was the case with children, some studies did not reveal a BA when administering this type of task to adults. For example, Emmorey et al. (2008) compared bimodal bilinguals (English-American Sign Language) (L2 AoO before 1) and unimodal bilinguals (English-other language) (L2 AoO during childhood) with English monolingual speakers in their forties and fifties. Results revealed a BA when completing the Flanker task only for unimodal bilinguals. This led the authors to propose that the positive effects of bilingualism on the EC could be somehow restricted to languages of the same modality. Kousaie and Phillips (2012b) put the BA hypothesis to the test with young adult university students (age range 18-35) using both behavioral data (RT and accuracy) and electrophysiological measures (Event Related Potentials –ERPs measures). While ERPs data revealed that English monolingual and English-French bilingual speakers (L2 AoO before 7) differed in "stimulus categorization and error-related processing" (Kousaie & Phillips, 2012b, p. 87), no differences between groups were detected on behavioral data. However, the authors highlight that the absence of a BA for behavioral data could be related to the type of participants, since the young adult population has been revealed as a particularly challenging population when it comes to detecting a BA effect (see Valian, 2015 or Vinerte, 2018 for a review).

Upon incorporating a methodology which implies specific manipulation of the tasks, different studies have highlighted that a BA among young adults manifests itself when complexity is added to the task. For instance, Costa et al. (2009) did not find better outcomes for Spanish-Catalan bilinguals (university students with L2 AoO during early childhood) when administering the ANT task in what the authors called "a low-monitoring version" (mainly one experimental condition presented), however, positive effects of bilingualism were observed when participants were administered what they refer to as "a high-monitoring version" of the task (congruent and incongruent experimental conditions mixed). The authors propose that in order to detect a BA among the young adult population it is necessary to present participants with especially challenging tasks, since otherwise such effects could go unnoticed. Similarly, Grundy et al. (2017) demonstrated that when performing the Flanker task young adult English monolingual speakers and English-other language bilingual speakers (university students with L2 AoO before 10) did not differ much; however, when taking into account the previous trial, bilingual speakers showed an advantage by experiencing less influence than monolingual speakers.

Unlike children and the adult population, one population that has received especially limited attention is adolescents. Little is known about the interaction between bilingualism and the EC abilities during the teenage age, but Chung-Fat-Yim et al. (2019) reported a BA for early-sequential English-other language bilinguals (L2 AoO before 7) in comparison to English monolingual speakers (mean age 16 years old) when completing the Flanker task.

Turning our focus now to the word-color Stroop task, it is important to note that, given its language component and its requirement of reading skills, this task has been said not to be very effective for studying children's EC abilities (Montgomery & Koeltzow, 2010). Therefore, it has been mainly used with adults and different Stroop-like tasks or similar suppression interference tasks have been proposed instead for younger individuals.

Esposito et al. (2013) compared the interference suppression abilities of Spanish-English bilinguals and English monolingual children between 3 and 5 years old. The authors considered bilinguals those children whose parents and teachers described them as fluent in both languages. Participants were administered a Color/shape task (see Esposito et al., 2013 for details) and results revealed better outcomes for bilinguals. These findings are in line with Martin-Rhee and Bialystok (2008), who also found that bilinguals outperformed monolinguals when the Simon task (Simon & Rudell, 1967) was administered to both English-French bilingual and English monolingual children (on average 5 years old). A positive effect of bilingualism to resolve interference has also been found with toddlers, since as Poulin-Dubois et al. (2011) showed, bilingual toddlers performed better in a shape Stroop task than their monolingual peers. However, Duñabeitia et al. (2014) administered a color-word Stroop task and a numerical Stroop task to Spanish-Basque bilinguals (L2 AoO during early childhood) and Spanish monolinguals (mean age range 8-13 years old) and did not detect better outcomes for bilingual children in either of the two experiments, which shows once more, that results are far from conclusive.

Focusing on the adult population, Bialystok et al. (2008) presented an English color-word Stroop task to two age groups of English monolingual and English-other language bilingual speakers: younger adults in their twenties and older adults in their sixties. Young adult bilinguals entered in contact with English before 6 and for the older adult bilinguals English was an L2. Accuracy data results revealed that monolinguals were less accurate than bilinguals and RT data showed an overall BA from which older bilinguals benefitted the most. These results were replicated in Bialystok et al. (2014), where participants with a similar profile to the previous study were administered a color-word paper Stroop task. Kousie and Phillips (2012a), on the other hand, did not detect differences between younger and older English monolinguals and English-French bilinguals (age range 18-35 for younger bilinguals and 60-84 for older bilinguals; L2 AoO before 8). The authors claim that their data do not show evidence of a BA since no differences between groups were detected for the incongruent condition. In fact, and even though the authors acknowledge "a general speed advantage" for bilinguals, as they called it, this is not interpreted in the study as an instance of a BA, although under the BEPA it would be. Ryskin et al.'s (2014) results also showed no differences between young adult monolingual and bilingual university students when administered the Stroop task and evaluating differences between groups in terms of conflict solving abilities.

In summary, previous research investigating the EC abilities of monolingual and bilingual speakers with the Flanker, Stroop or related tasks have evidenced conflicting results for both children and adults, revealing the young adult population as a particularly challenging one to detect a BA. Nonetheless, it is noteworthy that when an effect of bilingualism has been found this has been mostly on the positive side.

2.1.2. Metalinguistic abilities

The study of how bilingualism outlines the metalinguistic abilities of dual speakers has also attracted great attention within the literature. However, as Bialystok (2001) indicates, one of the main issues within the field has been, and still is, the lack of consensus about what researchers are referring to when using the metalinguistic term, since this term has been used "as a qualifier for at least three different entities: knowledge, ability and awareness" (Bialystok, 2001, p. 123). Following Adesope et al. (2010, p. 209), metalinguistic awareness has been traditionally defined as the ability to consciously reflect about language, and in this area, bilinguals –being language experts accustomed to dealing with two languages– are assumed to be more aware of the arbitrariness of language and of the different language idiosyncrasies. Bialystok also highlights the importance of the notion of attention, since "the combination of attention and explicit mental representations can produce the phenomenal experience awareness" (Bialystok, 2001, p. 126).

Previous research investigating the effects of bilingualism on the individuals' metalinguistic awareness abilities has revealed inconsistent results and it has been said that this inconsistency "begins to make sense when task demands are considered in terms of their connection to specific cognitive processes" (Bialystok, 2001, p. 135). Friesen and Bialystok (2012) carried out a detailed review of the effects of bilingualism on different metalinguistic tasks following Bialystok and Ryan's (1985) cognitive framework proposal to define metalinguistic abilities (henceforth referred to as Bialystok's metalinguistic cognitive framework). In this review, the authors highlight that a BA is usually detected in those tasks/measures that are challenging in terms of EC but that entail low demands of linguistic analysis, namely, the tasks/measures that are considered to be essentially metalinguistic. Conversely, no consistent BA has been found in those tasks/measures challenging in terms of linguistic analysis but that involve low demands of EC: the ones that are considered to be essentially linguistic task/measures.

Since the 80s, Bialystok and collaborators have tried to elucidate the interaction between bilingualism and metalinguistic abilities, and one of the tasks that they have repeatedly used is a grammatical judgment task (GJT), also referred to as a sentence judgment task, where grammar and semantics are manipulated. The number of experimental conditions can vary but there are always two fundamental experimental conditions: an essentially linguistic measure (ungrammatical but semantically sound sentences) and an intrinsically metalinguistic measure (grammatical but meaningless sentences). Participants are asked to determine the grammaticality of the sentence and, in order to succeed in those sentences including a semantic anomaly, participants need, on the one hand, to leave the semantic meaning aside (inhibition) and, on the other hand, to focus specifically on grammar (selective attention) (Bialystok, Peets, et al., 2014, p. 179).

Bialystok (1986b) examined the linguistic and metalinguistic abilities of English monolingual speakers in comparison to two different groups of bilinguals: English-other language bilingual children (other-language heritage speakers) vs. English-French bilingual children (immersion schooling) (age range 5-9 years old). The results showed no significant differences between groups when assessing linguistic knowledge but a significant BA when evaluating the metalinguistic abilities. Similar results were obtained in Bialystok and Majumder (1998), where the GJT was administered to English monolingual children and to English-French bilinguals (balanced bilinguals) and English-Bengali bilinguals (partially bilinguals)[11] (age range 7;11-9;11 years old). Again, bilingual speakers, irrespective of their bilingual background, outperformed monolinguals when determining the grammaticality of those sentences assessing the metalinguistic abilities but no differences between groups were detected for those sentences examining essentially linguistic knowledge. Bialystok et al. (2014) administered the GJT to anglophone children in grade 2 (age range 7-8.4 years old) and grade 5 (age range 9.9-11.5 years old) schooled either in English (monolingual speakers) or in French (becoming bilingual through immersion schooling). Their results mirrored those of previous research: no differences between groups in the linguistic measures but better metalinguistic abilities for bilingual speakers; however, this metalinguistic advantage was not detected until grade 5, which the authors interpret as a gradual development of the metalinguistic abilities. Cromdal (1999) administered a similar SJT to both partially and highly English-Swedish bilinguals and Swedish monolingual speakers (age range 6;4-7;3 years old). Results revealed that highly bilingual English-Swedish children outperformed monolingual speakers in both metalinguistic and linguistic abilities but no significant differences between Swedish monolinguals and partially Swedish-English were attested. Nonetheless, it is worth note that partial bilinguals scored higher than monolinguals in both measures.

The effects of bilingualism at the metalinguistic level have been mostly assessed with children, however some researchers have replicated the studies testing metalinguistic abilities with the adult population. For instance, Moreno et al. (2010) administered the GJT to English monolinguals and English-other language bilinguals (participants of immigrant origin who mostly acquired English before 12) (age range 19-28 years old) and collected both behavioral and ERP data. Results did not reflect different accuracy rates for the two groups, but ERP data suggested

[11] According to the authors, partial bilinguals refer to non-balanced bilinguals for which one of the two languages is stronger (Bialystok & Majumder, 1998; Poarch & Bialystok, 2015).

that “monolinguals did not process the semantic anomaly (no differences between correct and semantic anomalous mean amplitudes) but bilinguals were able to process it and achieve the same level of accuracy as monolinguals”. This is interpreted by the authors as being consistent with previous behavioral data focused on children (Moreno et al., 2010, p. 576). However, Paap and Liu (2014, p. 71) question this “counter intuitive” interpretation, as they called it, arguing that if bilinguals actually have better EC abilities this group should have been the one showing more impermeability to the semantic meaning and not the other way around. Paap and Liu (2014) replicated Moreno et al.’s (2010) study with behavioral data and did not found differences between English monolingual and English-other language adult bilingual university students. Both studies agree in that they do not show differences between groups at the behavioral level, thus mirroring the findings discussed in the previous section for the young adult population when performing the Flanker or the Stroop tasks.

Apart from the monolingual-bilingual comparison, the GJT has also been used to compare the linguistic and metalinguistic abilities of different types of bilinguals. Bialystok (1988) administered the GJT in both English and Italian to low English-Italian bilinguals and to high English-Italian bilingual children (age range 6.5-7 years old). All children were born and attended school in an anglophone context and Italian was the heritage language with different levels of proficiency. The results showed similar results for both versions of the task, therefore the authors ran the analysis combining the results of both tasks. Their findings revealed comparable metalinguistic abilities for both groups but better linguistic outcomes for the more experienced bilinguals. These linguistic and metalinguistic abilities have been tested beyond individual bilingualism and have also been assessed in children becoming bilingual through immersion school programs. Hermanto et al. (2012) administered this task in English and French to anglophone children enrolled in French immersion programs in grade 2 (mean age 7.7 years old) and grade 5 (mean age 10.6 years old). Their results showed that, overall, participants were more accurate when judging English sentences than when judging French sentences, but this gap became smaller with experience. Also, it was revealed that the higher difference between languages was found when judging ungrammatical but semantically sound sentences, which evidences that even though participants were being educated in French they were more sensitive to detect ungrammaticalities in English, their dominant language.

Similarly to the GJT, the WLJT[12] also entails high demands of EC, since, in order to succeed, participants need to put the semantic meaning aside (referent), and concentrate on the linguistic form (word length) (see Bialystok, 1986a). Therefore, this task mainly tests the arbitrariness that exists between a word form and its meaning (Cook, 1997). In order to shed light on whether bilingualism actually conferred an advantage in this respect, Bialystok (1986a) presented different word-pairs manipulating word length and object size to two groups of English monolingual children (group 1: mean age 4;10 year old; group 2: mean age 6;11 years old) and to an English-French bilingual group where participants had been enrolled in a French immersion programs for two years (mean age 6;10). The task included congruent items (longer word and bigger referent coincided) and incongruent items (longer word represented the smaller referent). Participants were asked to determine the bigger word (i.e., longest word) and the results revealed better outcomes for bilingual speakers. Nonetheless, this better outcome for bilinguals was not attested in Ricciardelli's (1992) study, where a similar task was administered to English-Italian bilinguals with different levels of proficiency (low and high) and to English monolinguals with different levels of proficiency too (age range 5-6 years old). In this task, different from Bialystok's

[12] Other terms that have been used within the field are word judgment task (Yelland et al., 1993), word length task (Ricciardelli, 1992) or size judgment task (Bialystok, 1986a).

(1986a), participants were presented with different words and they had to specify whether the word heard was a long or a short one independently of the referent size. Yelland et al.'s (1993) administered a similar task to English monolingual children and English-Italian "marginal bilingual" children, as the authors called them, of two age groups: preparatory grade (age range 4;4-5;6 years old) and grade 1 (age range 5;6-6;7 years old). Participants were tested at two time points: after two and after seven months of Italian instruction. In this case, children were presented with different images and were asked to name the object depicted and to determine whether that word was long or short. The marginal bilingual group included anglophone children who were receiving Italian lessons (one hour per week); thus, more than bilinguals, they seem to be L2 learners, as the authors also point out. Nevertheless, what is noteworthy is that, despite their limited exposure to an L2, the younger bilinguals showed better outcomes than those not receiving L2 instruction after only six months of L2 exposure. However, this BA was not detected among participants in grade 1, where both groups showed similar results. As Yelland et al. (1993) discuss, their findings contrast with Ricciardelli's (1992) and Cummins' (1976) defense of the threshold theory, according to which better metalinguistic outcomes would be observed only when a certain proficiency level has been achieved. Notwithstanding, it is important to highlight that although similar, Bialystok's, Ricciardelli's and Yelland et al.'s tasks are different. In all three, participants needed to block the semantic meaning in order to concentrate on the form and resolve the task. However, Ricciardelli's (1992) task was the easiest one of the three versions, since participants were only presented with one word and were asked to answer whether it was a long or a short word. Conversely, in Bialystok's (1986a) study, children were presented with a word pair instead of with a single word, which raised the level of complexity because participants not only needed to block one meaning but two in order to make the selection of the longest word. Similarly, in

Yelland et al.' (1993) experiment, before determining the word length, participants needed to first identify the image depicted, which also entailed lexical access. Taking all this into account, it could be the case that a BA would arise not only in those word-awareness metalinguistic tasks entailing high demands of EC but also in those tasks specially challenging from the data elicitation point of view, which would be somehow in line with Costa et al.'s (2009) proposal that some advantages are attested under certain experimental conditions. The WLJT has also been used with the adult population to study the metalinguistic abilities of illiterate speakers (Kolinsky et al., 1987) or to compare literate and illiterate individuals and, in these cases, results have shown better outcomes for the literate group (Kurvers et al., 2006).

In summary, previous research using a GJT to compare the linguistic and metalinguistic abilities of monolingual and bilingual speakers have revealed that the latter group, independently of the degree of bilingualism and context of acquisition, showed better metalinguistic abilities than their monolingual peers. Nonetheless, as was the case for the EC tasks presented in the previous section, this advantage, if present, seems to be easier to detect in children than in adults. We have also discussed the fact that results for essentially linguistic measures did not evidence a clear BA or disadvantage, supporting the idea that the results of these measures are somehow related to the participants' level of proficiency (i.e., vocabulary knowledge), as shown in Bialystok (1988). Regarding the effects of bilingualism in tasks similar to the WLJT included in this dissertation, the results were not clear-cut and it appears that the complexity of the experimental design might play an important role as previously proposed for the EC domain: the more cognitive challenging the task, the more bilinguals would display a BA.

2.1.3. Narrative abilities

The study of the narrative abilities has been extensively scrutinized among TD monolingual speakers, and to a lesser extent, among TD bilingual and multilingual speakers, even though during the last years there has been a growing body of research centered on this last group (see Berman & Slobin, 1994; Strömqvist & Verhoeven, 2004; Verhoeven & Strömqvist, 2001; among others). Narratives have been analyzed from different perspectives, but we are interested in achieving a better understanding of how monolingual and bilingual speakers compare in terms of macrostructural and microstructural abilities. As per Gagarina's (2016, p. 92) words "the macrostructure refers to the higher order hierarchical organization of the narrative text (…), while microstructure is represented by the internal, or language-specific, linguistic units used for the construction of coherent discourse". The bulk of bilingual research has mainly centered on how the macrostructural and microstructural abilities compare within and between languages (see Bitetti et al., 2020 for an overview). However, a direct comparison between monolingual and bilingual speakers with respect to these two abilities has received, comparatively, less attention, and the different studies that have addressed this comparison have revealed mixed findings, as has been the case with the EC and the metalinguistics results.[13]

Chen and Yan (2011) compared the narrative evaluative expressions of English-Chinese bilinguals (L2 AoO before 3) living in the US, to English monolinguals. The study included different age points: children of 5, 8 and 10 years old and young university students. Results revealed an effect of age, since the older the participants were, independently of whether they were monolingual or bilingual, the more evaluative clauses they included and the more a global narrative

[13] Given the high variability in nature and in number of the different macrostructural and microstructural measures included in the different studies reviewed in this section, we highlight the overall results and we refer the reader to the original papers for a detailed analysis of all the different measures included in each case.

evaluative perspective they took. Furthermore, even when exposed to English between 2-3 years later than monolinguals, bilinguals included a higher number of evaluative clauses. These results suggest a positive effect of bilingualism at the macrostructural level, though this is open to discussion, since some authors do not consider evaluative devices as instances of macrostructural abilities but as independent measures (see Hoang et al., 2018). Be that as it may, we need to be cautious with these results because as Chen and Yan (2011) point out, this BA could be confounded with a cultural effect due to the fact that previous research found a similar pattern for Chinese monolinguals when compared to English monolinguals. More recently, Andreou (2015) compared the macrostructural and microstructural abilities of Greek monolingual and Greek-other language bilingual children of two age groups (8-10 and 10-12 years old) (L2 AoO ranging from before 1 to after 6) using an oral telling task, an oral retelling task, and a written retelling task. At the microstructural level, monolingual speakers, both younger and older children, independently of the experimental tasks, outperformed bilinguals on producing longer and more syntactically complex narratives. At the macrostructural level, the pattern was reversed and bilinguals outperformed monolinguals at "capturing the global meaning of the story and [considering] the listener's perspective" (Andreou, 2015, p. 256). Likewise, younger bilinguals appeared to be better than their monolingual peers in the use of emotional and mental terms. Using a similar approach, Tsimpli et al. (2016) compared the macrostructural and microstructural abilities of Greek monolingual and Greek-other language bilingual children with and without DLD-SLI.[14] Participants were asked to retell a story and results showed different bilingual benefits for TD individuals for both microstructural and macrostructural abilities. At the microstructural level, no differences between groups were attested in terms of lexical diversity; however, better bilingual

[14] In this section, we only refer to the effects of bilingualism for TD individuals (mean age 9 years old); DLD-SLI results will be discussed in section 2.2.3.

outcomes were found when assessing children's subordination ratio. At the macrostructural level, groups did not differ on the use of more perceptual or physiological terms but bilinguals outperformed monolinguals in the use of more emotional and mental terms and by producing narratives with more complex story structures. Note that the three studies discussed up to now coincide in showing instances of a BA at the macrostructural level. However, Andreou (2015) and Tsimpli et al. (2016) differed regarding the effects of bilingualism at the microstructural level. This is not surprising if we take into consideration that microstructural measures are essentially a reflection of language proficiency (Andreou, 2015), and outcomes may vary depending on the participants' age (Rezzonico et al., 2016), language status (minority vs. majority) (Hipfner-Boucher et al., 2015), SES (Pearson, 2002), or school exposure and environment richness (Govindarajan & Paradis, 2019), to name a few factors.

Other studies, however, did not find instances of a potential BA advantage at the narrative level. For instance, Pearson (2002) compared the macrostructural and microstructural abilities of English-Spanish bilingual and English monolingual children living in the US of different age (age range 7-11 years old), SES, and schooling experience. Overall, results showed that monolinguals performed better but that the group differences lay principally at the microstructural level. Baldimtsi et al. (2016), following Andreou's (2015) and Tsimpli et al.'s (2016) methodological framework, compared Greek-other language bilingual children (TD and with ASD) to Greek monolingual children (TD and with ASD) (age range 7;3-11;6 years old). Focusing on TD children, the authors analyzed oral telling narratives at both the macrostructural and microstructural levels and here, unlike in Andreou's (2015) and Tsimpli et al.'s (2016) studies, bilinguals did not show better outcomes at either of the two narrative levels. Kunnari et al. (2016) also found similar macrostructural abilities between simultaneous Swedish-Finnish bilingual children and Finnish monolingual children (age range 5;0-6;6 years old) when analyzing their macrostructural abilities using both a telling and a retelling narrative task. Also, the bilinguals performed equally in both their languages. This outcome of similar results at the macrostructural level for both languages has been widely attested in the bilingual literature (Altman et al., 2016; Fiestas & Peña, 2004; Iluz-Cohen & Walters, 2012; among others) and is in line with Gagarina et al.'s (2016, p. 15) working hypothesis which states that narrative macrostructure, not being focused on the linguistic form, will be prone to crosslinguistic transfer due to its generalized rather than idiosyncratic features across languages. However, Kapalková et al. (2016), among others, have challenged this proposal, since they have revealed that early sequential Slovak-English bilingual children (age range 5-6 years old) exhibited better macrostructural outcomes in Slovak (L1) than in English (L2).

In a nutshell, previous findings, despite their different outcomes, coincide in showing that when effects of bilingualism are detected in the narrative process they are normally found at the macrostructural level and they present positive outcomes.

2.2. Bilingualism in non-typically developing populations

2.2.1. Executive control abilities

Previous research investigating the EC abilities of monolingual individuals with DLD-SLI and ASD has revealed dysfunctions in comparison to TD individuals (see Marton et al., 2019 for an overview of DLD-SLI studies and Nadig & Gonzalez-Barrero, 2019 for an overview of ASD studies). Bilingual research in this field is trying to elucidate whether, taking into consideration the cognitive characteristics of these individuals, bilingualism plays a role, and, if so, whether this is mainly positive as has been suggested for the TD population. Engel de Abreu et al. (2014) was one of the first studies exploring the EC abilities of bilinguals with DLD-SLI. In this study, the selective attention and interference suppression skills of early-sequential Portuguese-Luxembourgish (with and without DLD-SLI) bilingual children were compared to those of TD Portuguese monolingual children (age range 7;1-9;0 years old). Participants' selective attention was measured with the Sky search task (Mainly et al., 1999), a task where participants are required to identify identical paired images as fast as possible, and their interference suppression skills were assessed with the Flanker task. Results revealed that bilinguals with DLD-SLI did not significantly differ from their TD peers (monolinguals and bilinguals) on selective attention skills but that bilinguals with DLD-SLI underperformed TD bilinguals but showed comparable results to TD monolinguals when assessing their interference suppression abilities. It is premature to draw conclusions, but following Marton et al. (2019), these findings seem to suggest a potential positive effect of bilingualism in individuals with DLD-SLI. Laloi et al. (2017) also provide empirical evidence which confirms the existence of similar effects of bilingualism for DLD-SLI and TD populations. In this latter case, the authors investigated the response inhibition abilities of French-other language bilingual children with and without DLD-SLI (L2 AoO around 3) and French monolingual children with and without DLD-SLI (age range 6;0-8;6 years old). Participants were administered a visual Stop-signal task (Logan & Cowan, 1984) in which they were presented with a green or red arrow and were asked to determine its direction only when the arrow was green. No differences between monolingual and bilingual individuals with and without DLD-SLI were found, which is in line with previous studies comparing the response inhibition abilities of TD monolingual and bilingual individuals (Luk et al., 2010).

Gonzalez-Barredo and Nadig (2019b) analyzed the set-shifting abilities of monolingual and bilingual speakers (L2 AoO before 3 on average) with and without ASD.[15] Participants were

[15]All the studies reviewed in this section included ASD individuals without ID.

speakers of one or two of the following three languages: French, English-Spanish (age range 6-9 years old). The main objective of this study was to test whether the BA proposed for the TD population in relation to these EC abilities also held for the ASD population. To do so, the authors administered a Computerized Dimensional Change Card sort task (DCCS - Zelazo, 2006) where participants were presented with red rabbits and blue boat images and were asked to sort the images by color and by shape (counterbalanced among participants) at two different sorting stages by pressing two buttons labelled with a blue rabbit and a red boat. Only those participants who answered correctly five out of six trials for the second sorting dimension continued being tested. In this second part, whose difficulty level increased, they were instructed to sort the images by color if they had a border and by shape if they did not. Participants were considered to have succeeded in the task if they correctly classified 9 out of 12 trials. Their results confirmed the BAH, since the success rate for bilinguals with ASD was higher than that of monolingual speakers. Baldimtsi et al. (2016) also put the BAH to the test with Greek-other language bilingual children and Greek monolingual children with ASD (age range 7;7-11;6 years old) using both a Visual Attention global-local task (Navon, 1977) and an Updating n-back task (see Kane & Engle, 2002). In the first task, participants were presented with different shapes (crosses, circles, triangles, squares) in a big format made of the same shapes in a small format; manipulating congruency, participants were asked to determine either the large or the small shape. In the second task, participants were presented with different numbers and were asked to determine whether the number shown on screen coincided with the number they saw two trials back. Results revealed that bilinguals outperformed their monolingual peers in both tasks. These findings were replicated in Baldimtsi et al. (2020) and Peristeri et al. (2020) using similar tasks to those in Balditmsi et al.'s

(2016) and with the same participant profile: Greek-other language bilingual children with an age range between 6;9-15;6 years old.

In summary, the limited research available investigating the effects of bilingualism on the EC abilities rather than a negative effect seem to suggest a BA, as in the case of the TD research.

2.2.2. Metalinguistic abilities

The literature investigating the metalinguistic abilities of non-TD individuals is practically non-existent. However, there are two recent studies that follow this line of research. Peristeri et al. (2019) examined the metalinguistic abilities of Greek monolingual and Greek-other language simultaneous bilingual children with and without DLD-SLI (age range 8;4-13;6 years old) using a vocabulary subtest of the Wechsler Intelligence Scale for Children (WISC-III - Wechsler, 1991). In this experiment participants were presented with different words and were asked to define them. Thus, this metalinguistic task focused principally on linguistic knowledge without placing a burden on EC; nevertheless bilinguals with DLD-SLI outperformed their monolingual peers.

Gonzalez-Barrero and Nadig (2017) tested the metalinguistic abilities of monolingual and bilingual children with and without ASD (age range 5-10 years old). Participants were speakers of one or two of the following three languages: French, English, and Spanish, and were asked to produce the maximum number of words possible for different categories (animals and food in this specific case) within a minute. Although the authors analyze the number of words produced as instances of the lexical and semantic abilities of the ASD population, following Bialystok's metalinguistic cognitive framework, this task, as will be further discussed in chapter 3, could be considered as a metalinguistic one entailing a good command of vocabulary knowledge but low demands in terms of EC (Friesen & Bialystok, 2012). Results revealed that when monolinguals and bilinguals with ASD and with similar receptive vocabulary abilities are compared, bilinguals outperformed monolinguals by producing more correct words, which is interpreted by the authors as an instance of a BA in terms of "generativity or response initiation" (Gonzalez-Barrero & Nadig, 2017, p. 469).

These two studies show that bilingualism rather than having a negative effect on the metalinguistic abilities of these non-TD individuals, seem to have a positive one even when presenting participants with metalinguistic tasks assessing essentially linguistic knowledge.

2.2.3. Narrative abilities

The growing interest during the last years to better understand the narrative abilities of TD bilingual speakers has been extrapolated to the non-TD population. For some time now, narratives have been considered to be a good tool to assess bilinguals' language proficiency, since an evaluation of naturalistic data would be less biased than those derived from standardized tests (Cleave et al., 2010; Fiestas & Peña, 2004; Rojas & Iglesias, 2009; among others). Thus, the study of narratives has become a fertile locus for investigating how monolinguals with DLD-SLI and ASD compare to TD monolinguals first, and then, how bilinguals with DD diverge, if so, from the TD bilingual population and from their monolingual peers –although this last comparison has received the least attention (see Paradis & Govindarajan, 2018 for an overview). Cleave et al. (2010) was one of the first studies specifically comparing the macrostructural and microstructural narrative abilities of monolinguals and bilinguals with DLD-SLI. They recruited English monolingual children with DLD-SLI and English-other language bilingual children with DLD-SLI (mean age 4;4 years old). Both groups were administered a telling and a retelling task and results did not yield significant differences between groups either at the macrostructural or at the microstructural level. Focusing also on this population, similar results were found by Rezzonico et al. (2015) for DLD-SLI bilingual (English-other language) and monolingual (English) children, who were administered a retelling task twice with 6 a month difference (mean age 4;3 and 4;8 years old respectively). Results revealed similar macrostructural and microstructural abilities among DLD-SLI individuals, independently of their being monolingual or bilingual and a growth effect in both groups between both testing times. In a similar vein, Boerma et al. (2016), using a story telling task, revealed comparable macrostructural abilities for Dutch monolingual children and Dutch-other language bilingual children (L2 AoO before 4) with language impairment (age range 4;8-7;3 years old). The three studies reviewed displayed similar outcomes for both monolingual and bilingual speakers with DLD-SLI, which suggests that bilingualism does not affect their narrative skills negatively. Nonetheless, Tsimpli et al. (2016) found that Greek-other language bilingual children with DLD-SLI outperformed their Greek monolingual counterparts (mean age 9;2 years old) at some specific macrostructural measures using a retelling task. For instance, bilinguals mirrored monolinguals in the use of perceptual and physiological terms as well as in the use of emotional and mental terms but outperformed monolinguals by producing more structurally complex narratives. Conversely, when analysing microstructural abilities, bilinguals mirrored their monolingual peers in the use of subordination but showed shortcomings on the lexical diversity variable. This study, that also included TD monolingual and bilingual children, and whose results were discussed in section 2.1.3, showed that bilinguals with and without DLD-SLI had similar outcomes at the macrostructural level for the story structure variable. As Tsimpli et al. (2016) discuss, their results contrast with those of Squires et al. (2014), who found better macrostructural outcomes for TD bilingual children when compared to their language-impaired bilingual peers (age range 5-7 years old). Tsimpli and colleagues attribute the divergence of results between their study and Squires and colleagues' to a possible age factor effect, since their participants were older than those included in Squires et al. (2014). More specifically, they argue that older bilinguals with DLD-SLI would benefit from a higher exposure to their two languages and, as a consequence, this could redound somehow in a "greater ease to abstract shared properties of narrative macrostructure across their languages" (Tsimpli et al., 2016, p. 210). Although not discussed by Tsimpli and colleagues, this developmental hypothesis seems to also hold for explaining the BA they found at the macrostructural level, which differs from the previous research discussed in this section showing similar outcomes for monolingual and bilinguals with DLD-SLI. Tsimpli et al.'s participants were on average 4-5 years older than those included in Cleave et al.'s (2010) or Rezzonico et al.'s (2015) studies, and 2-4 years older than those in Boerma et al.'s (2016) study, which could somehow constitute a possible explanation for the lack of BA found in those studies. Interestingly enough, this same pattern seems to be present among individuals with ASD, although to a lesser extent.

Using a similar methodological framework to Tsimpli et al.'s (2016), Baldimtsi et al. (2016) elicited narratives from Greek monolingual children with ASD and Greek-other language bilingual children with ASD (mean age 9;8 years old) using a telling task. Their narrative abilities were also analyzed at the microstructural level and at the macrostructural level. While no differences between monolinguals and bilinguals with ASD were found at the microstructural level, bilinguals with ASD performed better than their monolingual peers at the story structure complexity level and when producing less referentially ambiguous pronouns. With a similar design, Peristeri et al. (2020) analyzed the narrative production of ASD Greek monolingual speakers and ASD Greek-other language bilinguals (mean age 9;7 years old) at the microstructural and macrostructural level and their results revealed a BA at both narrative levels. On the one hand, bilinguals produced more structurally complex narratives and used less ambiguous pronouns than monolinguals and, on the other hand, bilinguals produced more adverbial subordinate clauses than their monolingual peers but less complement clauses. In contrast to Baldmitsi et al. (2016) and Peristeri et al. (2020), Hoang et al. (2018) did not find differences between French monolingual and French-other language bilingual speakers (mean age 8;2 years old) with ASD neither at the microstructural level nor at the macrostructural level. Nonetheless, they did find differences between groups in relation to the number of utterances produced, which suggests a somehow higher productivity for bilingual speakers. The authors did not analyze narrative length as an instance of microstructural abilities but other authors have done so when differentiating microstructural productivity from microstructural complexity measures (Díez-Itza et al., 2018). If we consider narrative length in terms of microstructural abilities, Hoang et al.'s (2018) results would suggest certain BA at this level. On the contrary, and unlike the case in previous studies no positive effects at the macrostructural level were detected. Hoang et al. (2018) suggest that the differences found between their findings and Baldmitsi et al.'s (2016) and Tsimpli et al.'s (2016) results could be explained as a result of the differences in SES, narrative elicitation methods and microstructural and macrostructural measures as well as the differences in the participants characteristics. Also, taking into account Tsimpli et al.'s (2016) developmental proposal, it could be the case that age could be playing a role, since Hoang et al.'s (2018) participants were, on average, one year younger than those included in the previous studies showing a BA at the macrostructural level.

Taking into account the limited research available for DLD-SLI and ASD and the divergences between studies at different levels, as suggested by Hoang et al. (2018) (e.g., participants' profile, elicitation methods, analysis measures, etc.), it would be premature to draw conclusions. Nonetheless, it seems that a potential BA effect would lie more at the macrostructural level and when found, as was the case with the TD population, this effect would be positive.

2.3. Bilingualism in non-typically developing populations with genetic syndromes: an overview

As previously stated, individuals with genetic disorders have been traditionally discouraged from becoming bilingual. However, nowadays, we have evidence that individuals with genetic disorders such as Down syndrome (Feltmate & Kay-Raining Bird, 2008; Kay-Raining Bird et al., 2005), Williams syndrome (Barisnikov et al., 1996; Perovic & Lochet, 2015), Fragile X syndrome (Jarmendi, 2002) or Prader-Willi syndrome (García-Alcaraz, 2018; Liceras & García-Alcaraz, 2019), to name a few, do become bilingual. We also have evidence that the fact of speaking two languages does not seem to affect these individuals negatively (Cleave et al., 2014; among others). Nonetheless, as we describe below, the bulk of research investigating the effects of bilingualism in individuals with ID and genetic syndromes is still mainly restricted to the DS population, the most common chromosomal disorder entailing ID (Roberts et al., 2007).

The first studies centered on individuals with DS that were able to speak more than one language were carried out in the decade of the 90s and were mainly case studies aiming at describing the participants' cognitive and linguistic abilities. For example, Papagno and Vallar (1995) and Vallar and Papagno (1993) presented the case of EF, a young Italian woman in her twenties who in addition to Italian had knowledge of English and French, two languages to which she was exposed to during childhood while living abroad. Woll and Grove (1996) showed that the possibility of DS individuals becoming bilingual was not restricted to two languages of the same modality, since they presented the case of two hearing twins who were British Sign Language

(BSL) bimodal bilinguals (10 years old) who showed the same challenges in both their languages. English was the educational and social language with hearing individuals while BSL was the family language and the language used in social circles within the deaf community. Literacy skills have also been scrutinized using case study designs and the results have revealed that bilinguals with DS are perfectly able to read in two languages independently of whether or not both languages have the same alphabet, as is the case of Spanish and English (Nelson et al., 2008), or a different one, as in the case of the Russian-English combination (Burgoyne et al., 2016).

Apart from single case studies, the field has also tackled the effects of bilingualism among the DS population with group study designs, as in Kay-Raining Bird et al.'s (2005) pioneering study. In this study, the authors investigated the linguistic abilities of English-other language raised bilingual children with and without DS and the linguistic abilities of English monolinguals with and without DS (age range 2;0-11;4 years old). Participants were matched for developmental level and were asked to produce oral spontaneous data. They were also administered different standardized tests. Results revealed some shortcomings of the DS population in comparison to their TD peers but similar outcomes for bilinguals and monolinguals with DS, which the authors interpret as evidence that both groups were developing English linguistic abilities comparably and that bilingualism did not seem to play a negative role in the bilinguals' English development. In terms of the bilinguals' L2 abilities, the authors highlight variability and different degrees of bilingualism. However, it is suggested that DS individuals' L2 vocabulary acquisition mirrored that of the TD bilingual group. Similarly, Feltmate and Kay-Raining Bird (2008) investigated the vocabulary and morphosyntactic abilities of four English-French simultaneous bilingual children with DS. Each bilingual child was matched for developmental level with an English monolingual child with and without DS (age range 2;4-8;0 years old). Different standardized tests were administered, and an oral spontaneous sample was collected in which a comprehensive analysis of the participants' morphosyntactic abilities was performed. The results mirrored those of Kay-Raining Bird et al. (2005) since, overall, TD individuals outperformed individuals with DS but the two non-TD groups did not seem to differ much. In terms of differences between the bilinguals' two languages, the results replicated those obtained in the previous study: overall, better outcomes in English (mainly the dominant language) than in French. Better outcomes for the dominant than for the non-dominant language were also found by Trudeau et al. (2011). In this study the vocabulary acquisition development was scrutinized among English-other language (mostly simultaneous English dominant bilingual) children with DS. Participants' parents were asked to rate their children's receptive and expressive vocabulary abilities in both their languages. The results evidenced that participants had better vocabulary skills in English (dominant language) than in their L2, where there was more variability, and that their receptive vocabulary was stronger than their expressive vocabulary. Parents were asked to re-evaluate their children's vocabulary abilities at a later time and results suggested that while English vocabulary increased over time for 87% of the participants, only 60% showed progress on their L2. The authors discuss the importance of input for the L2 vocabulary acquisition process and argue for independence of development in both languages, since no negative effects of bilingualism were found on the acquisition process of the dominant language vocabulary. Cleave et al. (2014) analyzed the lexical learning abilities of simultaneous English-other language bilinguals and English monolinguals with and without DS (age range 3;0-19;3 years old). More precisely, the authors investigated the participants' fast mapping abilities by administering a syntactic bootstrapping task including both familiar and unfamiliar words of two different categories: verbs and nouns. The results showed that the four groups obtained better outcomes for the familiar condition and that for the unfamiliar condition participants performed better with nouns. Overall, TD bilinguals and DS bilinguals did not significantly differ while TD monolinguals performed better than monolinguals with DS. In terms of the effect of bilingualism, no differences were attested between monolingual and bilingual individuals within each group (TD and DS). This once more supports the claim that speaking more than one language does not seem to have a negative effect on the DS participants' dominant language development. More recently, Katsarou and Andreou (2019) have provided new evidence that no harmful effect of bilingualism was found when comparing the receptive and expressive linguistic abilities of Greek-English simultaneous bilinguals with DS, and Greek monolingual speakers with the same syndrome (age range 4-8 years old).

Even though the focus of previous research has mainly been the receptive and expressive linguistic abilities of bilinguals with DS, how bilinguals and monolinguals with DS compare at a more cognitive level has also been examined. For instance, Edgin et al. (2011) studied the effects of bilingualism, referred to as "second language exposure" by the authors, at the linguistic and cognitive level among English-Spanish bilinguals and English monolingual speakers with DS (age range 7-18 years old). To do so they administered a battery of different verbal and nonverbal tasks and their results did not reveal differences between groups either for the linguistic or for the cognitive measures. Similar results were found by Martin (2017), who compared the linguistic and EC abilities of one English-French bilingual speaker with DS (14 years old) attending a French immersion schooling program and a group of eight English monolingual speakers with DS (age range 12;1-17;9 years old). The participants' linguistic abilities were assessed using standardized measures and the EC abilities were examined by different experimental tasks putting the participants' inhibitory, cognitive flexibility and working memory abilities to the test. While the bilingual participant outperformed his/her monolingual peers at the linguistic level, he/she showed comparable EC abilities. Pinto-Cardona (2017) contrasted Spanish-English bilinguals with DS with English monolinguals with the same syndrome (age range 13-37 years old) in a series of tasks assessing the participants' short-term memory abilities (verbal, spatial and visual). The results revealed a significant BA for the visual short-term memory task, although in general bilinguals showed higher accuracy rates than their monolingual counterparts.

Although more research is needed, the repetitive outcome of not finding a negative effect of bilingualism for these individuals has opened the door to other lines of research unexplored until very recently, as is the case of the study of these individuals' foreign language acquisition abilities. For example, Polo (2017) investigated how four Spanish monolingual speakers with DS (age range 22-24 years old) learned English vocabulary in Madrid (Spain), an officially monolingual territory. The results revealed that participants were able to learn English vocabulary and that visual cues and an interactive approach among peers had a positive impact on the learning process.

Since its start, the study of bilingualism among individuals with genetic syndromes has gone through different stages which could be depicted as follows: in the early stages researchers were focused on proving whether it was possible for these individuals to become bilingual; then the focus was on discovering whether speaking two languages had a detrimental effect and, more recently, there is an interest to determine whether and how the proposed positive effects of bilingualism for the TD population are also found in these individuals. Given that no previous group study designs investigating the effects of bilingualism have been conducted with the PWS population, this dissertation aims to shed light on whether monolingual and bilingual individuals with PWS differ at the EC, metalinguistic and narrative level and if so whether bilingualism plays a negative or a positive role. The field, albeit growing, is still chiefly narrowed to the DS population and we will benefit from having more data exploring this issue with different genetic syndromes, different types of bilingualism, individuals at different developmental stages, and different language combinations.

3. Prader-Willi syndrome: an overview

This chapter presents an overview of the cognitive and linguistic profile of individuals with PWS. It is beyond the scope of this dissertation to provide an exhaustive description of the clinical and behavioral characteristics of this syndrome and we refer the reader to Butler et al. (2006) for a complete review. This chapter is divided in three sections. In section 3.1, we present a general outline of the cognitive an linguistic abilities of the PWS population; in section 3.2, we review the main findings of previous research examining the EC, metalinguistic and narrative abilities of monolingual speakers with PWS;[16] and finally, in section 3.3, we review the very few pieces of research dealing with the linguistic abilities of bilinguals with PWS that have been published. As was the case with the previous chapter, the revision of the different studies presented here focuses on the findings that are relevant for our specific purposes in spite of the fact that the scope of these studies is sometimes broader.

3.1. Cognitive and linguistic profile

PWS is a neurodevelopmental genetic disorder whose incidence has been estimated in one out of 20.000-25.000 births (Whittington & Holland, 2017). Despite its low incidence, it is considered to be one of the most common genetic syndromes among those identified and is not related to race, sex, SES or geographical origin (Alexander et al., 1995). PWS is caused by abnormalities in chromosome 15, more specifically by the "absence of expression of paternal genes from chromosome 15q11.2-q13" (Cassidy et al., 2012, p. 10). This region of chromosome 15 is the critical region for PWS and the lack of the paternal expression can obey, in Cassidy et

[16] All the studies reviewed in section 3.1 describe the cognitive and linguistic abilities of monolingual speakers with PWS and the majority of them did not include any reference to the participants linguistic background.

al.'s words (2012, p. 10), to "three primary mechanisms": Deletion (DEL), Uniparental disomy (UPD) and Imprinting mutation. DEL, which is the most common PWS subtype, affecting 65-75% of individuals, is caused by the absence of the paternal critical region of chromosome 15. UPD, the second subtype in incidence affecting 20-30% of the PWS population, is caused by the presence of two maternal copies of chromosome 15, and the Imprinting mutation subtype is caused by "a so-called imprinting defect", as defined by Cassidy et al. (2012, p. 10), which has a very low incidence as it only represents 1-3% of the cases (Cassidy et al., 2012; Foundation for Prader-Willi Research, n.d.).

The most distinctive trait of PWS is hyperphagia (extreme appetite), which, if not controlled, can cause morbid obesity (Cassidy & Driscoll, 2009). In fact, PWS is the principal genetic cause for obesity (Bueno Díez & Caixàs Pedragós, 2014) and these individuals' compulsive need to seek food is caused by the absence of sense of satiety due to an hypothalamus dysfunction (Cassidy & Driscoll, 2009). Other physical characteristics of individuals with PWS are hypotonia (poor muscle tone), short stature and distinctive facial features (Cassidy, 1997). They also present cognitive and developmental delays as well as behavioral disturbances (see Cassidy, 1997 for a complete review). For instance, as Cassidy et al. (2012) describe, individuals with PWS tend to be stubborn, controlling and/or manipulative and they usually experience temper tantrums as well as obsessive-compulsive behaviours.

It has been argued that the 15q11-q13 chromosome region is key for the ASD (Dimitropoulos et al., 2019; Dykens et al., 2011; Wang et al., 2018; among others). Thus, the relation between PWS and ASD has received considerable attention, especially during the last years. Although these two DD may share some social interaction difficulties, they display rather distinctive features. At first it was suggested that the prevalence of ASD among individuals with

PWS was higher than what it is actually believed to be, since according to Dimitropoulos et al. (2019, p. 4442), the incidence would be of 12.3% in contrast to the 25-41% previously proposed. However, UPD individuals seem to be more prone to autism because they show more similarities with the ASD individuals than those shown by their peers with DEL (Dimitropoulos et al., 2019; Dimitropoulos, Ho, et al., 2013; Dimitropoulos & Schultz, 2007; among others). In our participant sample (15 participants), two individuals had been diagnosed with UPD, two other individuals with Imprinting and no ASD confirmed diagnosis was disclosed for any of the participants of the study despite their belonging to either of the three groups.

Leaving aside the clinical studies, the bulk of previous research on PWS has mainly focused on these individuals' behavioral characteristics and, to a lesser extent, on their cognitive profile. The cognitive abilities of PWS individuals vary from person to person but, in general terms, they present mild to moderate ID with a full-scale IQ ranging on average between 50 and 70 (Dimitropoulos, Ferranti, et al., 2013; Dykens et al., 1992; Pegoraro et al., 2014; among others). There are also individuals that present an average IQ while other manifest a significant impairment (see Curfs et al., 1991). Nonetheless, unlike other genetic syndromes, individuals with PWS do not appear to experience an IQ decline over time (Dykens et al., 1992; Whitman & Thompson, 2006). Other cognitive areas that have been scrutinized and that seem to reveal certain strength are long-term memory in comparison to short-term memory (Conners et al., 2000), visual processing in contrast to auditory verbal processing skills (Curfs et al., 1991; Gabel et al., 1986), and simultaneous processing in comparison to sequential processing (Dykens et al., 1992). Besides this, PWS individuals have been found to outperform the TD population with jigsaw puzzle skills (Dykens, 2002) and individuals with DEL have been revealed to be specially good at it (Dykens, 2002; Veltman et al., 2004). However, it remains unclear whether these skills are due to a cognitive advantage or the result of mere practice (Whittington & Holland, 2017). In contrast to these strengths, PWS individuals show shortcomings with mathematical skills (Bertella et al., 2005) and they usually present deficits of attention (Gross-Tsur et al., 2001; Jauregui et al., 2007), a trait that has been carefully taken into consideration when designing the experimental tasks included in this dissertation. Differences between the IQ score of individuals with DEL and UPD have also been investigated and the findings are not necessarily straightforward. It has been proposed that these groups differ in terms of verbal IQ, with UPD individuals showing better outcomes (Roof et al., 2000). However Copet et al.'s (2010) and Dimitropoulos, Ferranti et al.'s (2013) results do not show differences between DEL and non-DEL individuals at the verbal IQ level, rather at the performance IQ level, favouring the DEL group. Additionally, these two studies revealed that in a within-group analysis non-DEL individuals showed better verbal IQ than performance IQ scores. It appears then, that in summary: DEL individuals score higher on performance IQ than non-DEL individuals; there is no difference between DEL and non-DEL individuals in verbal IQ; non-DEL individuals score higher on verbal IQ than performance IQ.

The study of the linguistic abilities of the PWS population has been rather limited (Kleppe et al., 1990) and, as a consequence, as Lewis (2006) points out, the picture of the linguistic profile of this population is not as "well" described as for other genetic syndromes that have traditionally received more attention, as is the case of DS, or WS, for example. However, what previous research has demonstrated is that the speech and language development of individuals with PWS are impaired, although the degree of impairment may vary from one person to another (Van Borsel et al., 2007). Among the different factors that may play a role in the individuals' variability, we find the level at which they are affected by different clinical features. For instance, as described in Lewis (2006) and Lewis et al. (2002), these individuals' mouth and dentition idiosyncrasies (e.g., narrow overjet and palatal arch) could affect their articulatory abilities and the hypotonia and larynx characteristic may play a role in these individuals' vocal pitch, among other vocal features. Similarly, Lewis and colleagues maintain that the PWS individuals' cognitive profile (e.g., mild to moderate ID) regulates, somehow, their receptive and expressive language abilities and the typical behavioural disturbances of this syndrome condition, as well as these individuals' social interaction and pragmatic abilities. In general, people with PWS show language delay and impairment with both receptive and expressive language (Lewis, 2006; Lewis et al., 2002); the latter having been said to be more affected (Branson, 1981; Downey & Knutson, 1995). Nevertheless, similar outcomes at both receptive and expressive levels have also been attested and even better expressive than receptive language abilities have been found for individuals with the UPD subtype (Dimitropoulos, Ferranti, et al., 2013). Likewise, it has been found that the linguistic abilities of the PWS population lag behind their chronological age (Kleppe et al., 1990; Van Borsel et al., 2007) and that their language abilities tend to be lower than what would actually be expected taking into account their cognitive abilities (Dimitropoulos, Ferranti, et al., 2013).

One of the linguistic areas that has received relative attention has been the study of these individuals' speech abilities (see Downey & Knutson, 1995; and Lewis, 2006 for overviews). Some distinctive features that have been documented are the presence of: (i) dysfluent speech (Defloor et al., 2000); (ii) acoustic and aerodynamic impairments (Defloor et al., 2001); (iii) articulation, intelligibility and voice deficiencies (Kleppe et al., 1990); and (iv) pitch and resonance problems (Åkefeldt et al., 1997). Morphosyntactic studies are scant but have revealed that the morphosyntactic development is delayed and that shortcomings in this respect are perceived since very early on (Atkin & Lorch, 2007). Downey and Knutson (1995, p. 145) argue that individuals with PWS show difficulties with morphology (e.g., plurals or past tenses) and that these shortfalls could be explained in terms of: (i) difficulties with morphological cues per se or (ii) challenges when producing certain sounds, which would be caused by articulatory problems. Conversational and spontaneous data have also revealed shorter Mean Length of Utterance (MLU) for PWS individuals in comparison to the TD population (Kleppe et al., 1990; Lewis et al., 2002; Van Borsel et al., 2007), which is interpreted as the presence of deficiencies in relation to the development of the morphosyntactic abilities. On the other hand, lexical variability abilities measured through the Type-Token Ratio (TTR) do not seem to be a particular challenge for this population (Van Borsel et al., 2007), nor does the process of reading (Cassidy, 1997; Dykens et al., 1992), and these two abilities are considered relative strengths for this population.[17] At the pragmatic level, according to Downey and Knutson (1995) and Lewis (2006), these individuals show difficulties in maintaining a topic, in keeping an appropriate physical distance with their interlocutor and in respecting conversational turns.

In summary, PWS individuals present behavioral disturbances as well as both cognitive and language deficits which affect their daily lives as well as their social interaction.

3.2. Executive control, metalinguistic and narrative abilities

The study of the EC control abilities of individuals with PWS has received certain degree of attention within the field and, with some exceptions (Walley & Donaldson, 2005), previous research has revealed systematic deficiencies of the PWS population in this respect. For example, Chevalère et al. (2013) analyzed the EC abilities of 20 individuals with PWS (age range 19-49 years old) to whom they administered the six subtests of the Behavioral Assessment of Dysexecutive Syndrome (BADS) (Wilson et al., 1996). Participants' results on their "planning,

[17] As Lewis (2006) points out, reading abilities refer to the reading decoding skills, since in some cases individuals show good reading decoding abilities but fail at reading comprehension.

organizing, switching/shifting, cognitive estimation, monitoring and inhibition" abilities (Chevalère et al., 2013, p. 312) were contrasted with those from the TD and, what the authors define as, the "neuropathological standardized population". Results revealed that the PWS individuals showed impaired EC abilities in comparison to the TD population but comparable skills to those described for the neuropathological group. Likewise, the authors also found a correlation between the full-scale IQ and the verbal IQ with the EC outcomes. One of the limitations of this work, as highlighted by the authors in a subsequent study (Chevalère et al., 2015), was the difficulty of drawing conclusions about specific EC deficits, since the different tasks administered evaluated diverse EC abilities at the same time. Thus, in an attempt to obtain a more accurate picture of the inhibition, switching and updating abilities of the PWS population, Chevalère et al. (2015) administered different computerized tasks to 17 TD individuals (age range 20-63 years old) and to 17 PWS individuals (age range 18-55 years old). As in the previous study, results revealed deficiencies for the PWS population in the three aforementioned EC abilities, whose performance seems to be globally related to the individuals' IQ but not in exclusivity. These results are in line with Woodcock et al. (2009), who compared the attention switching abilities of 28 TD individuals (age range 5;1-11;9 years old), 28 individuals with Fragile X syndrome (age range 9;2-19 years old) and 28 individuals with PWS (age range 6;10-18;7 years old) and found that both non-TD populations showed shortcomings and that in the specific case of the PWS individuals they persisted even when controlling for IQ.

Potential differences between the DEL and the UPD groups at the EC level have also been investigated. Stauder et al. (2005) investigated the inhibition abilities of 22 adult individuals with PWS (11 with DEL [mean age 26.7 years old] and 11 with UPD [mean age 27.7 years old]) and of 11 TD adults (mean age 27.3 years old). The study included both behavioral data (accuracy and

RT) and ERP data. The results showed that individuals with UPD significantly differed from DEL and from TD individuals by showing less accuracy and higher RT. At the neuronal level, both PWS subtypes showed deficits in terms of “early inhibition processes”, while at the “late general inhibition processes” only the UPD did (Stauder et al., 2005, p. 1468). Chevalère et al. (2013), when evaluating the EC abilities of the 20 participants with PWS previously presented, divided the sample in two groups: 14 individuals with DEL and 6 with UPD. Their results revealed that individuals with UPD performed worse than their DEL peers in two tasks measuring shifting and inhibition abilities, on the one hand, and planning, monitoring and organizing skills, on the other hand. These two studies coincide in showing certain advantage for DEL individuals at the EC level; nonetheless, results need to be taken cautiously given the low number of participants in general, and of individuals in the UPD group in particular.

To the best of our knowledge, no previous research explicitly addressing the metalinguistic abilities of the PWS population has been conducted. However, some studies intended to shed light on the EC of these individuals have administered different tasks that could be considered metalinguistic measures under the Bialystok’s metalinguistic cognitive framework. Friesen and Bialystok (2012) in their overview paper reviewed the results of different metalinguistic tasks among which they discussed two different fluency tasks: what they call a “category fluency task” and a “letter fluency task”. While in the former task participants are asked to produce the maximum words possible according to a semantic category (e.g., animals or objects) over one minute, in the latter they are asked to produce the maximum words possible starting with a specific letter (e.g., A or B) in the same amount of time. These two tasks, albeit similar, are said to tap different levels of EC, since as Friesen and Bialystok (2012) argue in the case of the letter fluency task, in order to succeed participants need to focus on the initial letter of the word while overlooking potential semantically related words that could come up as the result of the word produced. Jauregui et al. (2007) administered these two fluency tasks to 16 individuals with PWS (age range 17-48 years old) whose performance was compared to standardized scores from TD individuals and the results showed important shortcomings for the PWS population on both tasks. In addition, it was also revealed that the full IQ and the verbal IQ were significantly correlated with the results of both tasks. Walley and Donaldson (2005) also administered these two tasks to 18 individuals with PWS (age range 16-49 years old; 12 with DEL and 6 with UPD) and to 15 TD individuals (age range 18-49 years old) but, unlike Jauregui et al. (2007), these authors neither found differences between PWS and TD individuals nor between the DEL and the UPD subtypes.

Beyond the study of the linguistic abilities using standardized tests or very controlled experimental tasks (Dimitropoulos, Ferranti, et al., 2013; Van Borsel et al., 2007; among others), very few studies have investigated the linguistic abilities of this population using (semi)spontaneous data and those studies mainly focused on the analysis of these individuals' narrative abilities. Narrating is a complex process that implies the integration of both linguistic and non-linguistic knowledge (Tsimpli et al., 2016) and this has been proven to be challenging for the PWS population. Previous research investigating the narrative abilities of individuals with PWS has highlighted their poor narrative abilities and the impact that this fact has on their social relations and academic achievement (Lewis, 2006). Lewis et al. (2002) studied the narrative abilities of eight preschool children (age range 2-5 years old), eight school age children (age range 5-12 years old), and eight adolescent and adult individuals (more than 12 years old) with PWS. Participants were presented with a retelling task and their narratives were transcribed and analyzed according to the story grammar and the content items included. Participants were also asked to answer different inferential and factual questions about the story read. The results showed that the youngest participants had serious problems to perform both tasks (to retell the story and to answer the comprehension questions) and that while the narrative and comprehension abilities improved with age they continued to be strongly affected in the older participants. Garayzábel-Heinze et al. (2012) compared the narrative abilities of two individuals with PWS (12;9 and 13;9 years old), two individuals with WS (11;10 and 13;3 years old), and two individuals with Smith-Magenis syndrome (10;3 and 12;7 years old). Participants were presented with the picture book *Frog, where are you?* and were asked to narrate the story. Their narrative production was transcribed and analyzed according to three established measures: (i) *structure and coherence*, (ii) *process and complexity* and (iii) *content and multiplicity* (see Gonçalves, Henriques, & Cardoso, 2001; Gonçalves, Henriques, Alves, et al., 2001; Gonçalves, Henriques, Soares, et al., 2001 for an overview of this narrative protocol analysis). Results reflected that participants with the three genetic syndromes showed impaired narrative abilities. In the specific case of PWS, special difficulties were found for the structure and coherence and process and complexity measures, where they scored worse than their peers. In comparison to the previous two measures, PWS individuals showed relative strengths with the content and multiplicity one, since they showed similar outcomes to individuals with WS and slightly higher scores than individuals with Smith-Megenis syndrome. Despite using different elicitation methods, different narrative tasks and different target languages, both Lewis et al. (2002) and Garayzábal-Heinze et al. (2012) found parallel results: narrative shortcomings at the macrostructural level.[18]

[18] The authors did not discuss the linguistic profile of their participants. In Lewis et al.'s (2002) study we assumed that participants were English monolingual speakers because data collection took place in Ohio (USA), while Garayzábal-Heinze et al.'s (2012) participants were considered Spanish monolingual speakers as data collection took place in Madrid (Spain).

3.3. Bilingualism in Prader-Willi syndrome

Different clinical case study research papers have included bilingual participants. However, the bilingual abilities of these participants have been presented as a mere descriptive detail of the participants' characteristics and have not been discussed beyond making reference to their present or past ability to speak two languages (e.g., Buiting et al., 2000; Wey et al., 2005) if the bilingual exposure was disrupted. In other cases, the bilingual condition has been discussed as a possible explanation for the unexpected results obtained. For example, Nugnes et al. (2013) analyzed the cognitive and behavioral profile of an eight-year-old sequential bilingual Spanish-Italian child with PWS (UPD). Focusing on the cognitive profile results of this individual, the authors studied his full, verbal and performance IQ. The findings revealed that he had a borderline full-scale IQ and that the performance IQ scores were higher than the verbal IQ outcomes. These results were not in line with previous findings revealing that PWS individuals with UPD showed better results for the verbal than for the performance IQ (see section 3.1). The authors attributed this unexpected finding to a possible effect of bilingualism, since it is inferred that the participant was tested in his/her non-dominant language. In this case, although bilingualism was an intrinsic characteristic of the participant, the study did not specifically focus on elucidating how speaking two languages might affect this participant's IQ.

To the best of our knowledge, no previous research specifically addressing the study of bilingualism among the PWS population has been conducted beyond García-Alcaraz (2018) and Liceras and García-Alcaraz (2019). Both studies are case studies including an English-Spanish male adult bilingual with PWS. Spanish was this individual's heritage language and English the environmental and dominant language. García-Alcaraz (2018) compared this individuals' (33 years old at the time of testing) narrative abilities to those of a TD English-Spanish adult bilingual with a similar linguistic background. They were both presented with two picture books from the Frog Stories series and were asked to narrate each picture book in both languages in different experimental sessions. Narratives were analyzed at the macrostructural and microstructural level. Microstructural analysis focused on the MLU and the TTR indexes while the macrostructural analysis, as in Garayzábal-Heinze et al. (2012), followed Gonçalves and collaborators' evaluation protocol (Gonçalves, Henriques, & Cardoso, 2001; Gonçalves, Henriques, Alves, et al., 2001; Gonçalves, Henriques, Soares, et al., 2001). Results revealed that the narrative abilities of the PWS individual were compromised at both the macrostructural and the microstructural level but that the shortcomings detected were comparable in both English and Spanish. Also, the MLU and the macrostructural abilities improved the second time the participant with PWS narrated each story.

Liceras and García-Alcaraz (2019), for their part, investigated this individual's representation of grammatical gender (34 years old at the time of testing) using codeswitched structures. As has been suggested, the use of codeswitched structures allows one to determine whether speakers of two languages differing with respect to the presence or absence of grammatical gender transfer the gender of the gendered language to the genderless one (Liceras et al., 2016). This phenomenon, known as Analogical Criterion (AC) (Otheguy & Lapidus, 2005) has been found to be respected by TD bilinguals whose dominant language has grammatical gender. Namely, these individuals favor switches involving a gender matching determiner + noun switch as in (1) versus the non-matching version in (2). They also favor the gender matching noun–adjective sequence, in (3) over the non-matching version in (4) (see Liceras et al., 2016 for an overview and a detailed discussion of similar examples). Thus these individuals seem to apply the AC in both structures when performing a SJT in which they are asked to rate different sentences as in (1-(4, or in a Sentence Completion Task (SCT), where they are asked to provide either the determiner of the adjective when asked to complete phrases and sentences such as those listed in (1) to (4) along the lines of Liceras et al. (2016).

(1) La$_{[theF]}$ ventana$_{[windowF]}$.

(2) El$_{[theM]}$ ventana$_{[windowF]}$.

(3)The ventana$_{[la\ ventanaF]}$ is blanca$_{[is\ whiteF]}$.

(4)The ventana$_{[la\ ventanaF]}$ is blanco$_{[is\ whiteM]}$.

The English-Spanish bilingual with PWS was administered both the SJT and the SCT designed and used by Liceras et al. (2016) with TD Spanish-English bilingual children and adults (either Spanish or English dominant). Results showed that the PWS participant, despite being English dominant, performed similarly to the TD Spanish dominant bilinguals described in the aforementioned study in the sense of abiding by the AC, especially in the SCT.

These two studies, with all due caution given their case study nature, rather than a detrimental effect of bilingualism, show a positive effect and are in line with previous research centered on investigating bilingualism in individuals with other genetic syndromes (see section 2.3 for a review).

In broad terms, in this chapter it has become apparent that the PWS population show deficiencies at both the cognitive and the linguistic level and that these individuals show impairments with EC, metalinguistic and narrative abilities. The next step is to determine whether, within their limitations, monolingual and bilinguals with PWS show comparable abilities and whether bilingual individuals may benefit from bilingualism similarly to what has been proposed for the TD population (see section 2.1) and more recently for some individuals with DLD-SLI and ASD (see section 2.2).

4. Overview of the study

In this chapter we introduce the various aspects of the study conducted to investigate whether and how monolingual and bilingual individuals with PWS compare in terms of linguistic and cognitive abilities and whether or not Spanish-Catalan bilinguals show instances of a BA similarly to what some studies have found for TD individuals, and, more recently, for non-TD individuals. The main objective of this chapter is to present the reader with an outline of the general study which will set the foundation for the detailed analyses of the participants' EC (chapter 5), metalinguistic (chapter 6), and narrative (chapter 7) abilities. We first present the research questions together with their corresponding hypotheses (4.1). Next, we present the standardized tests administered (4.2.), as well as a detailed description of the participants' characteristics and their linguistic and cognitive profile (4.3). We then provide a general overview of the different tasks administered (4.4), in addition to the specification of the procedure followed for conducting the data collection (4.5). The chapter finishes with a general description of the statistical analyses carried out (4.6).

4.1. Research questions and hypotheses

The specific research questions that this dissertation is set up to answer and the corresponding hypotheses that will be tested are as follows:

Research question 1: *How do bilinguals and monolinguals with PWS compare with respect to EC, in particular in relation to the interference suppression abilities?*

Hypothesis 1: based on previous findings concerning the effects of bilingualism on individuals with genetic syndromes (see section 2.3), we do not anticipate a bilingual detrimental effect for the PWS population. If a bilingual effect is detected and it is somehow similar to that described for the TD population (see section 2.1.1) and, more recently, for non-TD individuals with ASD (see section 2.2.1), this effect is expected to be positive. More precisely, taking into account Hilchey and Klein's (2011) review of results for the TD population, a potential BA would manifest as a general processing advantage (BEPA hypothesis) rather than as an smaller interference effect (BICA hypothesis).

Research question 2: *How do bilinguals and monolinguals with PWS compare with respect to metalinguistic abilities?*

Hypothesis 2: as we anticipated in the previous research question, no negative effects of bilingualism are expected for the PWS population. Likewise, if bilingualism plays a similar role in the metalinguistic abilities of both TD and non-TD individuals, following Bialystok's metalinguistic cognitive framework (Bialystok & Ryan, 1985), we expect that dual speakers with PWS will outperform their monolingual peers on challenging measures in terms of EC (i.e., when judging grammatical but meaningless sentences or when determining the longest word of a pair where the word size and the word length are incongruent). Conversely, no BA is anticipated when evaluating measures with low demands of EC (i.e., when judging ungrammatical but meaningful sentences, which assesses the participants' grammatical knowledge; and when determining the longest word of a pair, where the word size and the word length are congruent). In fact, in the specific case of the SJT, if differences between monolinguals and bilinguals are detected when judging ungrammatical but meaningful sentences, these are expected to favor Spanish monolinguals, since bilinguals are predominantly Catalan dominant.

Research question 3: *How do bilinguals and monolinguals with PWS compare with respect to narrative abilities at the macrostructural and microstructural levels?*

Hypothesis 3: as in the previous two research questions, bilingual individuals with PWS are not expected to show shortcomings when their narrative abilities are compared to those of their monolingual peers. If an effect of bilingualism is similar between the TD and non-TD population (see sections 2.1.3 and 2.2.3) at this level, a potential BA may arise when comparing their macrostructural abilities. However, due to the fact that our bilinguals are predominantly Catalan dominant, no BA is anticipated at the microstructural level. In this case, if a between-group difference is detected, monolinguals are expected to show better outcomes.

In addition, and even though they do not constitute core aspects of this dissertation, we are also interested in analyzing the global effect of the participants' individual differences in terms of nonverbal IQ, receptive vocabulary, and sentence recall abilities on the results obtained, as well as in investigating whether bilinguals' metalinguistic and narrative abilities differ in Spanish and Catalan. Our fourth and fifth research questions address these two additional objectives as follows.

Research question 4: *Do participants' individual differences in terms of nonverbal IQ, receptive vocabulary, and sentence recall abilities play a role in explaining some of the overall data variance observed when analyzing the participants' EC, metalinguistic, and narrative ability results?*

Hypothesis 4: previous research investigating the EC and metalinguistic abilities of the PWS population has suggested that these individuals' IQ may play a role in their performance (see section 3.2). Taking this into account, it may be the case that despite the lack of overall significant group differences, individual differences according to nonverbal IQ, as well as to receptive vocabulary and sentence recall abilities could have an impact on the participants' EC, metalinguistic and narrative abilities (RT, accuracy rates and narrative assessment).

Research question 5: *Do bilingual speakers show comparable metalinguistic and narrative abilities in both Spanish and Catalan or do they perform better in Catalan, the predominantly dominant language?*

Hypothesis 5: the relevance of the language dominance factor among TD individuals has been extensively discussed within the literature (see Treffers-Daller, 2019 for an overview). Anticipating a similar effect among non-TD individuals, bilinguals with PWS are expected to perform better in Catalan than in Spanish, particularly when dealing with measures that evaluate principally linguistic knowledge (i.e., narrative abilities at the microstructural level and essentially grammatical judgments in the SJT).

4.2. Questionnaires and standardized tests administered

In order to have a better understanding of the participants' family profile and of their language fluency background, the participants' parents were asked to complete two questionnaires created for that purpose.[19] Additionally, the participants' nonverbal IQ, receptive vocabulary (also an indirect measure of verbal IQ and proficiency level), and sentence recall abilities (also an indirect measure of expressive language and working memory capacities) were assessed using

[19] The family profile and language fluency background questionnaires used in this dissertation were an adaptation of the documents used for that purpose at the Language Acquisition Research Laboratory of the University of Ottawa.

different standardized tests. Monolingual speakers were administered the receptive vocabulary and the sentence recall tests in Spanish while bilinguals completed both tests in Spanish and Catalan.[20] All the standardized tests were administered by the author of this dissertation, who is a Spanish-Catalan bilingual and who was trained on how to administer those tests before data collection took place. To have an accurate picture of the cognitive and linguistic profile of our participants is crucial, because, unlike in the TD population, and as already discussed in chapter 3, for PWS individuals there is normally a gap between their chronological age and their language development (Kleppe et al., 1990; Van Borsel et al., 2007). Furthermore, it has been shown that these individuals' linguistic abilities tend to be lower than what could actually be expected when considering their cognitive abilities (Dimitropoulos, Ferranti, et al., 2013).

In the next section we provide a succinct description of the standardized tests used, as well as of the family profile and the language fluency background questionnaires completed by the participants' parents.

Family profile background questionnaire

Throughout this questionnaire, participants' parents provided information about their languages and their level of study (primary, secondary, postsecondary and university), as well as the participants' general language knowledge and social background information, i.e., gender, age or PWS subtype. Following Calvo and Bialystok (2014, p. 281), the family SES was determined according to the parents' educational level, since, as these authors and other

[20] As will be further discussed in detail in this section, there are no Catalan standardized versions available to assess the participants' receptive vocabulary and sentences recall abilities. Consequently, translations/adaptations of the Spanish and English versions of these tests were used to elicit data intended to test these linguistic abilities in Catalan.

researchers suggest, this variable normally correlates with other socioeconomic variables (e.g., career and salary).

Language use and fluency questionnaire

Participants' parents also completed a language use and fluency questionnaire where they specified when the participants' started to speak, read and write; the language of instruction at different developmental levels; the Spanish and/or Catalan percentage of language used with different interlocutors, and the effort required in Spanish and/or Catalan to carry out different linguistic activities involving oral and written comprehension and production skills. The measure taken into account to compare the participants' effort for oral/written comprehension/production skills was the mean score (using a scale from 1 to 9; 1 being very easy and 9 very difficult) computed from the different tasks included for each linguistic skill (oral/written comprehension/production).

Nonverbal intelligence test

Participants' nonverbal IQ was assessed through the *Test of Nonverbal Intelligence 2* (TONI-2) (L. Brown et al., 1990), a test where no language interaction is involved and which assesses the nonverbal cognitive development of individuals between 5;0-85;11 years old. In this test participants are presented with different abstract problem-solving items of increasing difficulty.

Receptive Vocabulary test

Participants' receptive vocabulary abilities were evaluated with the standardized Spanish version of the *Peabody Picture Vocabulary Test* (PPVT-III) (Dunn et al., 2006), which is meant to be administered to individuals between 2;5-90 years old. As a standardized version of the PPVT is not available in Catalan, the form A of the PPVT-4 English version (Dunn & Dunn, 2007) was translated/adapted into Catalan and administered to dual speakers.[21] In both Spanish and Catalan tests participants were orally presented with different words of increasing difficulty and were asked to select the picture (out of 4) that they considered best depicted the word presented. The PPVT has been proposed to be a good proxy for mental age discrimination (Garayzábal-Heinze et al., 2012) and a good indicator of language proficiency level (Bialystok, Peets, et al., 2014). Previous research has also revealed a high correlation between the PPVT and the verbal IQ for TD children (Hodapp & Gerken, 1999), TD adults (Bell et al., 2001) and non-TD individuals with ASD (Krasileva et al., 2017). Consequently, the PPVT seems to be a solid measure that goes beyond its primary function of evaluating the receptive vocabulary (see Strauss et al., 2006, pp. 940–953 for a complete review of the PPVT and its implications).

Sentence recall test

The sentence recall test administered was the standardized subtest of the Spanish version of the *Clinical Evaluation of Language Fundamentals-CELF 5* (Wiig et al., 2018). Participants were presented orally with different sentences of increasing complexity in terms of length and language and were asked to repeat the sentences exactly as heard without introducing any changes. This task, beyond evaluating individuals' expressive language, has also been argued to be a good indicator of the individuals' working memory abilities (Alloway & Ledwon, 2014). Although the Spanish version of this test is conceived for individuals between 5-15 years old, all the participants'

[21] To translate an standardized version of the PPVT to a language that does not have one seems to be standard practice within the field (Bialystok & Barac, 2012, p. 61; Cleave et al., 2014, p. 46; among others). Different from the English materials, the Spanish PPVT version only includes one form. Thus, in order to avoid administering the Spanish test twice, with all the implications that this might have for the results reliability, we decided to resort to the English version. Consequently, Catalan results are not standardized and need to be interpreted cautiously.

sentence recall abilities were evaluated using this version even if their chronological age was above this cut-off. In those cases, following Dimitropoulos et al. (2013), who used the English version of the CELF subtest with PWS individuals over its upper age limit (21 years old), we calculated the participants' standardized measures as if they were 15 years old. As in the case of the PPVT, there is not a standardized version of this test in Catalan. Therefore, the sentences were translated and adapted into Catalan and were administered in successive days. Once more, Catalan results are not standardized and results need to be interpreted prudently.

4.3. Participants' cognitive and linguistic profile

A total of 16 participants (eight bilinguals and eight monolinguals) were recruited in Spain. Monolingual speakers were recruited and tested in the Autonomous Community of Madrid (Spanish monolingual region) and bilingual speakers were recruited and tested in Catalonia (Spanish-Catalan bilingual region). One bilingual participant was excluded from the study because he/she did not complete all the tasks and the tasks completed did not contain enough material to be evaluated. Thus, the final sample consisted of 15 participants: eight monolinguals (M = 22.80 years old, SD = 11.58, age range 9;4-47;0) and seven bilinguals (M = 18.04 years old, SD = 8.95, age range 10;5-33;10), of which six were female and nine were males. Females were evenly distributed across groups. All participants had been diagnosed with PWS and no other medical condition potentially affecting their cognitive or linguistic abilities was disclosed. The sample included participants with different PWS subtypes (DEL n = 11, UPD n = 2 and Imprinting n = 2). Participants with Imprinting were evenly distributed across groups and the two individuals with UPD were bilingual speakers. All participants completed all the experimental and narrative tasks, however, in specific cases one or several participants were excluded from a particular data sample.[22]

In terms of schooling, participants had different educational experiences. In Spain the education competencies are transferred to the different autonomous communities, hence, taking into account that monolingual and bilinguals were from two different regions (Madrid and Catalonia), the two groups had different schooling experiences. Even within the same autonomous community the participants' educational characteristics might have varied since each educational center had different policies to integrate students with special needs. However, despite the possible differences that may exist, what is important to highlight is that the majority of them have attended inclusive primary and secondary school with curriculum adaptation. Nevertheless, as each individual is unique, in certain cases the schooling experience has been primarily that of the special education system or a combination of both an inclusive and special education schooling experience. The language of education in Madrid is Spanish while in Catalonia it is Catalan. However, a recent study shows that although Catalan is the language of education in Catalonia, Spanish seems to be the most used language in the playground in both primary and secondary schools among Catalan students in urban areas (see Plataforma per la llengua, 2019), where our bilingual speakers were recruited. Moreover, despite the fact that education is imparted in Catalan, the Spanish proficiency level of Catalan students does not seem to be compromised. Between 2009 and 2010, the National Institute of Educational Evaluation (INEE Spanish acronym), under the Spanish Ministry of Education, carried out the General Diagnosing Surveys to assess, among other

[22] One bilingual was excluded from the participants' sample in the Stroop task due to colour blindness and one monolingual and one bilingual were excluded in the narrative task because of unintelligibility and/or distraction. When this happened, the essential group characteristics for that specific data sample were provided in the participants' section in chapters 5 and 7.

capacities, the Spanish proficiency level of students in fourth and eighth grade in all the national territory. Results revealed that the Spanish proficiency level of Catalan students was not only within the national average but exceeded that of some students receiving education only in Spanish in different monolingual autonomous communities (INEE, 2010, 2011). Independently of whether they were educated in Madrid or in Catalonia, most of the participants were receiving or had received instruction in a foreign language during their school period, English being the most common option, although in one case it was French. In the case of monolingual speakers, Spanish was the environmental and educational language and the only language used within the family and social circles.[23] For bilingual speakers, Catalan and Spanish were both environmental languages but Catalan was the language of education and the main language of communication within the family and social circles.[24] According to the bilingual parents' responses in the language use and fluency questionnaire, the participants' use of Spanish (in percentage) with the following interlocutors was as follows: mother (M = 22.86, SD = 33.52, range 0-80); father (M = 32.86, SD = 44.99, range 0-100), caregiver (M = 40, SD = 52.92, range 0-100); friends in school (M = 31.43, SD = 33.38, range 0-90); friends outside of school (M = 32.86, SD = 28.70, range 0-80); coworkers (M = 45.00, SD = 35.36, range 20-70).[25] We can therefore conclude that our bilingual participants are Catalan dominant with the exception of one participant, who, despite coming from an eminently Catalan speaking family, the family kept Spanish as the main language of communication with him/her as he/she received special education in Spanish and they were recommended to focus on one language. Following Montrul (2008), our participant sample could

[23] One monolingual participant lived and attended typical schooling (grade 1) in The United Kingdom during one year. Currently this participant only speaks Spanish and is considered by his/her family a monolingual Spanish speaker.
[24] In the case of one bilingual speaker the father's first language was Serbian. Although the bilingual participant is able to recognize some basic words in Serbian the family considers him/her as only Spanish-Catalan bilingual.
[25] The mean calculation of the language spoken with the caregiver and the coworkers was only for the participants who declared to have linguistic interchanges with these people (three cases for each case).

be defined as early-sequential bilinguals, since for the majority of them Catalan was the family language and their formal contact with Spanish started in kindergarten or primary school when they were on average between 3-6 years old.[26] Therefore, from this moment onwards the bilingual group will be referred to as Catalan dominant early-sequential bilinguals, since this is the label that best represents the idiosyncrasies of this group, a group that we have to acknowledge as being quite homogenous given the idiosyncrasies of the population. Interestingly enough, as will be discussed later in this section, even though Spanish is supposed to be the "weaker" language for bilinguals, according to the standardized tests results, their competence in Spanish does not seem to differ from the one observed among monolingual speakers. A Catalan monolingual control group is not included in the study because Catalan monolingual speakers are virtually nonexistent (see Boix-Fuster & Sanz, 2008).

Parametric tests were used to analyze the participants' demographic, cognitive and linguistic characteristics. Norman (2010, p. 631) defends that "parametric statistics can be used with Likert data, with small sample sizes, with unequal variances, and with non-normal distributions, with no fear of 'coming to the wrong conclusion'". Additionally, the use of parametric methods seems to be the norm within the field even when including small samples (Cleave et al., 2014; Edgin et al., 2011; Kay-raining Bird et al., 2005; among others). Pearson product-moment correlation (also called Pearson's *r*), *t*-tests, repeated-measures analysis of variance (ANOVA) and mixed-design ANOVAs were used to analyze participants' characteristics. Post hoc pairwise comparisons using the Bonferroni correction were conducted when ANOVAs yielded significant results. The alpha level was set at 0.05.[27] Effect sizes for *t*-tests were estimated using Cohen's

[26] For one bilingual, Spanish was the family language and the formal contact with Catalan started at school. However, this participant nowadays identifies himself/herself as a Catalan dominant bilingual and the family corroborates this information.
[27] This alpha level applies to all the statistical analyses presented in this dissertation.

index (d) [28] and effect sizes for ANOVAs were estimated using partial eta squared (η_p^2).[29] Both SPSS (version 25) (IBM Corp., 2018) and JASP (version 0.11.1) (JASP Team, 2019) statistical software were used to carry out the statistical analysis and compute the effect sizes. Sphericity was checked with the Mauchly's test and if this assumption was not met, the violation was reported and the degrees of freedom were adjusted using the Greenhouse-Geisser correction ($\varepsilon < .75$) or the Huynh-Feld correction ($\varepsilon > .75$) (Field, 2009, p. 461).[30]

Table 1. Chronological age and parental educational level by group.

	Chronological age		Maternal Education[a]		Paternal Education[a]	
	Monolinguals	**Bilinguals**	**Monolinguals**	**Bilinguals**	**Monolinguals**	**Bilinguals**
M	22.80	18.04	3.38	3.86	3.43	3.83
SD	11.58	8.95	0.92	0.38	0.79	0.41
Range	9;4-47;0	10;5-33;10	2.00-4.00	3.00-4.00	2.00-4.00	3.00-4.00

[a]Maternal and parental education range: 1 = primary school, 2 = secondary school, 3 = postsecondary school, 4 = university studies.

The participants' sample ranged from children to adults, adolescents and young adults being the profile with highest incidence (see appendix A for a description of the participants' characteristics). We are aware of the wide age range included in our sample but to recruit participants with PWS, especially bilingual speakers, is not an easy task. Nonetheless, despite the age diversity included in the sample (see Table 1) the two groups did not differ significantly in terms of overall chronological age ($t(13) = 0.88$, $p = .395$, $d = 0.46$). As regards to SES (measured through parental level of education), results showed a strong correlation between maternal and parental education ($r(13) = .82$, $p < .001$). Therefore, the maternal education variable was selected as the SES variable of reference because, unlike the parental education variable, it did not contain

[28] The proposed interpretation for d is the following: 0.2 (small); 0.5 (medium) and 0.8 (large) (Cohen, 1988).
[29] The proposed interpretation for η_p^2 is the following: 0.01 (small); 0.06 (medium) and 0.14 (large) (see Richardson, 2011).
[30] This procedure was also applied in successive analyses.

missing values. Results did not display differences between groups in terms of SES ($t(13) = -1.30$, $p = .218$, $d = -0.67$); hence, both monolinguals and bilinguals seem to share a common SES background: middle class families.

Table 2. Speaking, reading and writing onset age by group.

		Bilingual		Monolingual
		Catalan	**Spanish**	**Spanish**
Speaking Onset Age	*Mean*	3.00	4.14	2.81
	SD	1.04	1.60	1.16
	Range	1.50-4.50	1.50-6.00	1.50-5.00
Reading Onset Age	*Mean*	6.64	7.07	6.19
	SD	2.53	2.32	1.51
	Range	5.00-12.00	5.00-12.00	4.00-8.00
Writing Onset Age	*Mean*	6.43	6.86	6.31
	SD	2.51	2.34	1.44
	Range	5.00-12.00	5.00-12.00	4.00-8.00

To have a better understanding of the participants' onset age for speaking, reading and writing in Spanish (see Table 2), a mixed-design ANOVA with Skill (Speaking, Reading and Writing) as the within-subjects variable and Group (Monolingual vs. Bilingual) as the between-subjects factor was run. The Mauchly's test indicated that the sphericity assumption was not meet for the repeated measure variable ($\chi 2(2) = 27.05$, $p < .001$), thus, degrees of freedom were adjusted using the Greenhouse-Geisser correction ($\varepsilon = .53$). Results revealed a main effect of Skill ($F(1.06, 13.72) = 24.81$, $p < .001$, $\eta_p^2 = 0.66$) but not of Group ($F(1, 13) = 1.78$, $p = .205$, $\eta_p^2 = 0.12$). The interaction between both factors was not significant ($F(1.06, 13.72) = 0.30$, $p = .608$, $\eta_p^2 = 0.02$). Pairwise comparisons with the Bonferroni correction showed a significant mean age difference between the reading onset age and the speaking onset age ($p < .001$, $d = 1.34$) and between the speaking onset age and the writing onset age ($p < .001$, $d = -1.35$) but not between the reading onset age and the writing onset age (p = 1.000, d = 0.07). These results show that for both monolingual and bilingual speakers the speaking skill develops earlier than reading or writing, which is expected, and that the reading and writing learning processes go hand in hand when participants are in primary school.

A repeated-measures ANOVA with Skill (Speaking, Reading and Writing) and Language (Spanish and Catalan) as the within-subjects variables for bilingual speakers revealed a main effect of Skill ($F(1.07, 6.44) = 10.64$, $p = .015$, $\eta_p{}^2 = 0.64$) but not of Language ($F(1, 6) = 2.90$, $p = .140$, $\eta_p{}^2 = 0.33$) and the interaction between both factors was not significant ($F(2, 12) = 2.89$, $p = .095$, $\eta_p{}^2 = 0.33$). As the Mauchly's test indicated a violation of the sphericity assumption for the main effect of Skill ($\chi 2(2) = 9.91$, $p = .007$), its degrees of freedom were adjusted using the Greenhouse-Geisser correction ($\varepsilon = .54$). The results indicate that the speaking, reading and writing onset age of bilinguals does not significantly differ in both languages.

To analyze the parental evaluation of the receptive/expressive oral/written abilities in Spanish (see **Table 3**), a mixed-design ANOVA with Mode (Oral and Written) and Process (Comprehension and Production) as the within-subjects variables and Group (Monolingual vs. Bilingual) as the between-subjects factor was carried out. Results showed a main effect of Mode ($F(1, 13) = 38.05$, $p < .001$, $\eta_p{}^2 = 0.75$) but not of Process ($F(1, 13) = 0.57$, $p = .469$, $\eta_p{}^2 = 0.04$) or Group ($F(1,13) = 1.52$, $p = .240$, $\eta_p{}^2 = 0.11$). None of the interactions was significant: Mode × Group (F(1, 13) = 0.67, $p = .427$, $\eta_p{}^2 = 0.05$), Process × Group ($F(1, 13) = 1.83$, $p = .199$, $\eta_p{}^2 = 0.12$) and Mode × Process × Group ($F(1, 13) = 3.34$, $p = .091$, $\eta_p{}^2 = 0.20$). These results suggest that for both monolinguals and bilinguals written tasks entail a higher difficulty than oral tasks independently of the process involved: comprehension/ production.

Table 3. Parental evaluation of oral/written comprehension/production abilities by group.

		Bilingual		Monolingual
		Catalan	**Spanish**	**Spanish**
Oral Comprehension	*Mean*	3.17	3.94	3.53
	SD	1.85	1.34	0.67
	Range	1.38-6.38	1.88-6.13	2.50-4.50
Oral Production	*Mean*	3.09	3.78	3.54
	SD	1.21	0.74	1.44
	Range	1.46-4.71	2.58-4.92	2.29-6.00
Written Comprehension	*Mean*	5.04	5.76	4.11
	SD	1.47	0.55	1.61
	Range	3.13-7.43	5.17-6.86	1.50-6.43
Written Production	*Mean*	5.30	5.47	5.65
	SD	2.18	1.94	1.22
	Minimum	2.80-8.00	2.80-8.00	4.00-7.33

Oral/Written comprehension/production evaluation range from 1 (very easy) to 9 (very difficult).

To evaluate potential differences between Spanish and Catalan for bilingual speakers (see **Table 3**), a repeated-measures ANOVA with Mode (Oral and Written), Process (Comprehension and Production) and Language (Spanish and Catalan) as the within-subjects variables was run. Results showed a main effect of Mode ($F(1, 6) = 25.05$, $p = .002$, $\eta_p^2 = 0.81$) and a significant interaction between Mode and Language ($F(1, 6) = 13.16$, $p = .011$, $\eta_p^2 = 0.69$). No main effect of Process ($F(1, 6) = 0.01$, $p = .913$, $\eta_p^2 = 0.00$) nor Language ($F(1, 6) = 1.88$, $p = .220$, $\eta_p^2 = 0.24$) were detected as any of their interactions: Mode × Process ($F(1, 6) = 0.02$, $p = .890$, $\eta_p^2 = 0.00$), Process × Language ($F(1, 6) = 2.02$, $p = .205$, $\eta_p^2 = 0.25$), Mode × Process × Language ($F(1, 6) = 4.88$, $p = .069$, $\eta_p^2 = 0.45$). Post hoc pairwise comparisons with the Bonferroni correction for the interaction between Mode and Language yielded significant mean differences between Catalan oral skills and Catalan written skills ($p = .010$, $d = -2.02$); between Spanish oral skills and written Spanish skills ($p = .021$, $d = -1.74$); and between Catalan oral skills and Spanish written skills (*p*

= .006, *d* = -1.64). These results are in line with the previous analysis revealing that written tasks are harder than oral tasks in both Catalan and Spanish.

Table 4. Receptive vocabulary, sentence recall and nonverbal IQ measures by group.

		Bilingual		Monolingual
		Catalan	**Spanish**	**Spanish**
Receptive Vocabulary Standard Score[a]	*Mean*	91.57	67.86	64.13
	SD	18.01	14.50	16.95
	Range	67.00-118.00	55.00-91.00	55.00-94.00
Receptive Vocabulary Raw Score	*Mean*	169.57	110.71	117.88
	SD	26.74	24.36	25.33
	Range	134.00-198.00	87.00-157.00	86.00-162.00
Receptive Vocabulary Age Aquivalent	*Mean*	12.93	9.48	10.18
	SD	4.19	3.19	3.30
	Range	8;5-18;11	7;5-16;3	7;5-17;2
Sentence Recall Standard Score	*Mean*	71.43	72.14	66.88
	SD	19.73	20.59	14.87
	Range	55.00-105.00	55.00-105.00	55.00-100.00
Sentence Recall Raw Score	*Mean*	30.14	29.29	28.13
	SD	21.76	23.60	15.55
	Range	5.00-66.00	4.00-64.00	11.00-62.00
Sentence Recall Age Equivalent[b]	*Mean*	6.76	6.89	6.18
	SD	4.45	4.88	3.82
	Range	3;0-15;10	3;0-15;10	3;0-15;10
			Bilinguals	**Monolinguals**
Nonverbal IQ Standard Score	*Mean*		90.14	75.75
	SD		11.01	10.35
	Range		79.00-112.00	55.00-87.00
Nonverbal IQ Raw Score	*Mean*		18.71	12.75
	SD		4.79	6.16
	Range		14.00-25.00	2.00-22.00

[a] IQ measures are normally presented in standard scores (*M* =100, *SD* 15). Therefore, this score is used to make results easily interpretable and comparable among measures.
[b] The CELF-5 interpretation manual does not provide specific chronological age equivalents below or above 3;00-15;10 but the following interpretations: ≤3;00 or ≥ 15;10 years old. Thus these two age points are used.

A mixed-design ANOVA with Standardized test (Nonverbal IQ, Receptive Vocabulary, and Sentence Recall) as the within-subjects variable and Group (Monolingual vs. Bilingual) as the between-subjects factor revealed a main effect of Standardized test ($F(2, 26) = 5.32$, $p = .012$, $\eta_p^2 = 0.29$) but not of Group ($F(1, 13) = 2.96$, $p = .109$, $\eta_p^2 = 0.19$). The interaction between both factors was not significant ($F(2, 26) = .551$, $p = .583$, $\eta_p^2 = 0.04$). Post hoc pairwise comparisons with the Bonferroni correction showed a significant mean difference between nonverbal IQ and receptive vocabulary abilities ($p = .015$, $d = 0.86$) (see **Table 4** for descriptive data). These results do not indicate differences between groups with respect to these three standardized measures, but they do reveal better nonverbal IQ abilities than receptive vocabulary abilities for both monolingual and bilingual speakers. We will come back to this issue at the end of this section.

No correlations were detected between the three standardized measures (standard scores): sentence recall, and nonverbal IQ ($r(15) = -.09$, $p = .742$); sentence recall and receptive vocabulary ($r(15) = .20$, $p = .471$) and nonverbal IQ and receptive vocabulary ($r(15) = .06$, $p = .840$). Likewise, no strong correlation was found between the standard scores of the three standardized tests and the chronological age variable (see descriptive data in **Table 4**): chronological age and nonverbal IQ ($r(15) = -.08$, $p = .778$); chronological age and receptive vocabulary ($r(15) = -.39$, $p = .152$) and chronological age and sentence recall ($r(15) = .08$, $p = .768$). This is not surprising if we take into account that chronological age is essential to computing the standard score and the three measures were hence adjusted by it. Additional analysis comparing the raw scores (not adjusted by age) of the three standardized tests and the chronological age did not reveal a strong correlation, which supports the previously mentioned gap between chronological age and developmental age for PWS individuals: chronological age and nonverbal IQ ($r(15) = .10$, $p = .731$); chronological age and receptive vocabulary ($r(15)= .45$, $p = .092$) and chronological age and sentence recall ($r(15) = .11$, $p = .708$).

The standard scores of the three standardized tests were analyzed next with gender, SES and PWS subtype as the grouping variable. A mixed-design ANOVA with Standardized Test (Nonverbal IQ, Receptive Vocabulary and Sentence Recall) as the within-subjects variable and Gender (Male vs. Female) as the between-subjects factor revealed a main effect of Standardized test ($F(2, 26) = 6.47$, $p = .005$, $\eta_p{}^2 = 0.33$) but not of Gender ($F(1, 13) = 0.15$, $p = .708$, $\eta_p{}^2 = 0.01$). The interaction between both factors was not significant ($F(2, 26) = 1.82$, $p = .182$, $\eta_p{}^2 = 0.12$). The mixed-design ANOVA with Standardized test as the within-subjects variable and SES (Secondary, Postsecondary and University Education) as the between-subjects factor showed a marginal significant main effect of Standardized test ($F(2, 24) = 2.87$, $p = .077$, $\eta_p{}^2 = 0.19$) but not of SES ($F(2, 12) = 0.96$, $p = .410$, $\eta_p{}^2 = 0.14$). The interaction between both factors was not significant ($F(2, 24) = 0.340$, $p = .848$, $\eta_p{}^2 = 0.05$). Thus, no differences between gender and SES were attested. The mixed-design ANOVA with Standardized test as the within-subjects variable and PWS subtype (DEL, UPD, Imprinting) as the between-subjects factor yielded a main effect of Standardized test ($F(2, 24) = 4.65$, $p = .020$, $\eta_p{}^2 = 0.28$) and a marginal significant main effect of PWS subtype ($F(2, 12) = 3.47$, $p = .065$, $\eta_p{}^2 = 0.37$). The interaction between both factors was significant ($F(4, 24) = 3.471$, $p = .023$, $\eta_p{}^2 = 0.37$). Pairwise comparisons with the Bonferroni correction for the significant interaction reveled the following significant mean differences: nonverbal IQ (DEL) and sentence recall (DEL) ($p = .017$, $d = 1.04$); sentence recall (DEL) and sentence recall (Imprinting) ($p = .007$, $d = -1.08$); and sentence recall (Imprinting) and receptive vocabulary (DEL) ($p = .029$, $d = 0.94$). It is premature at this point to draw conclusions about the role of the PWS subtype taking into account the low variation across participants (more than 70%

of the participants belonged to the DEL group). However, and even if we are looking at the data from a rather exploratory stand, these results suggest that when comparing subtypes, individuals with Imprinting appear to show better sentence recall abilities than individuals with DEL.

A repeated-measures ANOVA with Test (Receptive Vocabulary and Sentence Recall) and Language (Spanish and Catalan) as the within-subjects variables for bilingual speakers yielded a significant main effect of Language ($F(1, 6) = 30.60$, $p = .001$, $\eta_p^2 = 0.84$) but not for Test ($F(1, 6) = 0.72$, $p = .430$, $\eta_p^2 = 0.11$). Nevertheless, the interaction between both factors was significant ($F(1, 6) = 50.97$, $p < .001$, $\eta_p^2 = 0.90$). Overall, results show a higher marginal mean for Catalan than for Spanish (see **Table 4** for descriptive data), which is not surprising taking into account that bilingual speakers are predominantly Catalan dominant. Post hoc pairwise comparisons with the Bonferroni correction for the Test × Language interaction showed better receptive vocabulary abilities in Catalan than in Spanish ($p < .001$, $d = 3.33$). On the other hand, bilinguals demonstrated comparable sentence recall abilities between languages ($p = 1.000$, $d = -0.10$). These results could evidence that more than an expressive language measure the sentence recall test may be more a working memory one, since no differences among languages were attested, unlike with the receptive vocabulary measure. Given that Catalan tests are not standardized, these results need to be interpreted cautiously.

In a nutshell, these results show that both groups are similar in terms of nonverbal IQ and, despite bilingual speakers being Catalan dominant, both groups show comparable receptive vocabulary and sentence recall abilities in Spanish.

As previously discussed, the standardized test results suggest participants' deficiencies in both verbal (measured trough the PPVT) and nonverbal IQ, with nonverbal IQ being higher than the verbal IQ. Given that the majority of participants pertain to the DEL group, these results are in line with previous studies investigating the intellectual functioning of the PWS subtype with a higher incidence (Copet et al., 2010; Dimitropoulos, Ferranti, et al., 2013). For many years, scoring below 70 (standard score) in IQ standardized tests was the standard diagnosis criteria for ID. These days, however, a combination of both clinical assessment and standardized tests results are used to estimate an individual's intellectual functioning (American Association on Intellectual and Developmental Disabilities, n.d.). According to the American Association on Intellectual and Developmental Disabilities (AAIDD) (2020) "intellectual disability is a disability characterized by significant limitations in both intellectual functioning and in adaptive behavior, which covers many everyday social and practical skills". In a similar vein, the American Psychiatric Association (APA) (2020) maintains that "intellectual disability is identified by problems in both intellectual and adaptive functioning". Whitman & Greenswag (1995, p. 127) support this comprehensive view when they argue that it could be the case that individuals with PWS obtain high IQ scores but behave nonetheless as individuals with mild ID. Thus, taking into account the fact that monolingual and bilingual participants were neither assessed with different IQ battery tests nor was their adaptative functioning evaluated by a clinician, we cannot conclude that they present ID. This being said, for the purpose of this dissertation, a general approach is taken considering all participants as individuals with ID, since this is not only suggested by their standardized tests results but also by the fact that behavioral disturbances and mild to moderate intellectual disabilities are two characteristics, among many others, of the PWS population (see chapter 3).

4.4. Materials: general overview of the experimental tasks administered

Four experiments were designed and administered in order to elicit experimental and spontaneous data to address the five research questions outlined in 4.1. In this section we do not present an exhaustive description of each task, as this is included in the experimental design sections in the subsequent chapters, but we provide the reader with a glimpse of the different experiments included. Participants' EC abilities were scrutinized trough two experimental tasks: the Flanker task, and the Stroop task (experiment 1). The participants' metalinguistic abilities were analyzed by means of two offline tasks: a Sentence Judgment Task (SJT-experiment 2), and a Word-Length Judgment Task (WLJT-experiment 3). Finally, the participants' narrative abilities were examined at the macrostructural and the microstructural level using an oral narrative elicitation task (experiment 4).

Experiment 1: EC experiment

Two modified versions of the original Flanker (Eriksen & Eriksen, 1974) and Stroop tasks (Stroop, 1935) were created. While in the former task, participants were presented with different sequences of five chevrons and were asked to determine the direction of the central chevron, in the latter, participants were presented with different color-colored words and were asked to determine the font color. In both tasks, congruency was manipulated (congruent vs. incongruent) and a neutral baseline condition was also included. Two dependent variables were analyzed: RT and response accuracy.

Experiment 2: Sentence Judgment Task (SJT)

A modified version of the extensively used SJT by Bialystok (Bialystok, 1987, 1992; among others) was created and adapted to Spanish and Catalan. Participants were orally presented with different sentences and were asked to determine the grammaticality of the sentence heard, independently of meaning. Four experimental conditions were included manipulating grammar and semantics. Grammaticality was manipulated according to a linguistic dimension variable:

order vs. morphology. The dependent variable for this task was the participants' response accuracy.

Experiment 3: Word-Length Judgment Task (WLJT)

A modified version of the WLJT-type task used in Bialystok (1986a), Ricciardelli (1992) and Yelland et al. (1993) was created and adapted to Spanish and Catalan. Participants were orally presented with different word-pairs and were asked to determine the longest word of the pair. Four experimental conditions were included manipulating word length and referent size. For each experimental condition half of the items were animate and the other half were inanimate. The dependent variable was the participants' response accuracy.

Experiment 4: Narrative Task(s)

Monolingual and bilingual speakers were asked to narrate the picture book *A boy, a dog and a frog* (Mayer, 1967) in Spanish. Bilinguals were also asked to narrate the picture book *Frog, where are you?* (Mayer, 1969) in Catalan. Narratives were evaluated at the macrostructural level using the Narrative Scoring System (NSS) and at the microstructural level by means of the MLU and TTR indexes. The dependent variables were the holistic and analytic narrative macrostructural scores as well as the MLU and TTR scores.

4.5. Data collection procedure

Data collection took place in Spain in both Madrid and Barcelona. All participants expressed their willingness to freely participate in the study by completing and signing a consent form. Additionally, one of the parents gave his/her consent for them to participate in the study by signing a different consent form. The study met the ethical standards of the University of Ottawa Research Ethics Board and was approved under the certificate number S-08-18-722.

Table 5. Experimental sessions for data collection.

Experimental session 1	Experimental session 2	Experimental session 3
Bilingual speakers		
1. Receptive vocabulary test — Spanish (20 minutes) 2. Sentence recall test — Spanish (10 minutes) 3. Test of nonverbal IQ (20 minutes) 4. Narrative A boy, *a dog and a frog*—Spanish (15 minutes)	1. Word-length judgment task —Spanish (15 minutes) 2. Executive control task (15 minutes) 3. Sentence judgment task —Spanish (15 minutes) 4. Narrative *Frog, where are you?*—Catalan (15 minutes)	1. Receptive vocabulary Test —Catalan (20 minutes) 2. Sentence recall test —Catalan (10 minutes) 3. Word-length judgment task —Catalan (15 minutes). 4. Sentence judgment task —Catalan (15 minutes)
Monolingual speakers		
1. Receptive vocabulary test – Spanish (20 minutes) 2. Sentence recall test – Spanish (10 minutes) 3. Test of nonverbal IQ (20 minutes) 4. Narrative A boy, *a dog and a frog*—Spanish (15 minutes)	1. Word-length judgment task — Spanish (15 minutes) 2. Executive control task (15 minutes) 3. Sentence judgment task —Spanish (15 minutes)	
Approximate time: 55 min	*Approximate time: 60 min*	*Approximate time: 60 min*

As shown in Table 5, data collection took place in three experimental sessions for bilingual participants and in two experimental sessions for monolingual participants. Each session had an approximate duration of one hour. During the first session both groups were administered the three standardized tests in the following order: receptive vocabulary test, sentence recall test (both in Spanish) and the nonverbal IQ test. The first experimental session finished with the narration of the picture book *A boy, a dog and a frog* in Spanish. During experimental session 2 participants completed, in the following order, the WLJT, the EC experiment and the SJT. Both metalinguistic tasks were administered in Spanish and the language-based EC task (Stroop task) was completed in the participants' dominant language.[31] Bilingual speakers also narrated the picture book *Frog, where are you?* in Catalan. During experimental session 3 bilingual speakers completed the receptive vocabulary test, the sentence recall test, and the two metalinguistic tasks (WLJT and AJT) in Catalan.

4.6. Data analysis overview

In this section, we present an overview of the different statistical analyses carried out within the framework of this dissertation. We first present the between-group analyses followed by the within-group analyses for bilingual speakers.

In order to determine whether monolingual and bilingual individuals with PWS show comparable EC, metalinguistic, and narrative abilities we performed both regression analyses and mixed-effects models analyses.[32] Regression analyses were used in those cases where each participant contributed one data point, such as when analyzing the MLU, the TTR or the holistic narrative evaluation. Mixed-effects models analyses, on the other hand, were used when participants contributed multiple data points in a repeated measures design, as in the analytic narrative evaluation or in the EC and metalinguistic tasks.

---

[31] Previous studies show that language dominance plays a role in RT when performing the Stroop task (Marian et al., 2013). Therefore, in order to avoid this effect, participants completed the Stroop task in their dominant language.

[32] Mixed-effects models analyses are top-level powerful and flexible statistical analyses that include both fixed and random effects, which accounts for their name. In simple terms, fixed effects are the variables expected to predict the data (i.e., independent variables or predictors) and random effects are those effects "that can be expected to have a non-systematic, idiosyncratic, unpredictable, or 'random' influence in your data" (Winter, 2013, p. 39), as, for example, by-participant or by-item variation (see Winter 2013, 2019 for a complete review). Also, this type of analysis has been claimed to be robust even when facing missing data (Linck & Cunnings, 2015) and to be a solid alternative to ANOVA analyses due to, among other factors, its better "handling of categorical data" (Barr et al., 2013, p. 256). Mixed-effects models analyses are gaining strength in our area and are being increasingly used within the field of bilingualism and L2 acquisition with both large and small samples (for examples of studies including 10 ≤ participants per group see Ivanova et al., 2014; Kartushina & Martin, 2019; Van der Linden et al., 2018, among others).

Regression analyses and mixed-effects models analyses were run in R (version 3.6.2; R Core Team, 2019) using the RStudio interface (version 1.2.5042; RStudio Team, 2019) with the *lme4* (version 1.1-23; Bates et al., 2015) package installed. The continuous dependent variables[33] were modeled using either the regression analysis with the *lm* function (i.e., MLU, TTR and holistic narrative evaluation) or the linear mixed model analysis with the *lmer* function (RT data or the analytic narrative evaluation). The binary dependent variables (i.e., accuracy in terms of correct or incorrect, such as the accuracy dependent variables in the EC and metalinguistic tasks) were modeled using a generalized linear mixed model analysis with the *glmer* function. Mixed-effects models analyses were carried out taking into account all data points instead of the by-participant mean per condition that is common in ANOVA analysis.

In total, as illustrated in Table 6, this dissertation includes three regression analyses (*lm* models) and seven mixed-effects models analyses (three *lmer* models and four *glmer* models). We focus first on the modelling process of the mixed-effects models.

Table 6. Continuous and binary dependent variables modelled with lm, lmer and glmer functions.

Continuous dependent variables Multiple regression analysis (*lm*)	Continuous dependent variables Mixed-effects models (*lmer*)	Binary dependent variable Mixed-effects models (*glmer*)
1. Mean Length of Utterance data	1. Flanker task RT data	1. Flanker task accuracy data
2. Type Token Ratio data	2. Stroop task RT data	2. Stroop task accuracy data
3. Holistic narrative evaluation	3. Analytic narrative evaluation	3. Sentence Judgment Task accuracy data
		4. Word-Length Judgment Task accuracy data

In order to avoid potential problems due to overfitting (models too complex for the available data) and/or collinearity (high correlation between variables), which could lead to

33 Results derived from Likert Scale evaluations, such as in the holistic or analytical narrative evaluation, were analyzed as continuous variables (see Kizach, 2014; Norman, 2010 for further details).

unreliable results, non-convergent models, or singularity fit issues –among other problems– the mixed-effects models were kept as simple as possible taking into account the study design and the data available (see Frost, 2017, 2019 for more details about how to avoid common problems in regression analysis in general and overfitting in particular). To that end, the analyses included both Experimental Condition and Group, as well as their interaction, as fixed effects. Additionally, in both metalinguistic tasks, a second repeated measures variable was included in the models (Linguistic Dimension for the SJT and Animacy for the WLJT). In these cases, interactions between Group and the two repeated measures variables were restricted to two-way interactions in an attempt to not overfit the models. To reduce collinearity, following the standard practice within the field (de Bruin et al., 2018; Linck & Cunnings, 2015; among others), all continuous predictors were *z*-scored and all categorical variables with two levels (i.e., Group, Linguistic Dimension and Animacy) were contrast coded (-0.5 vs. 0.5). The *nloptwrap2* optimizer (see Bolker, 2014) was used to enable the models' convergence (*nloptr* -version 1.2.2.1- package; (S. G. Johnson, n.d.).

The modelling process was carried out in two phases (phase 1 and phase 2). In phase 1, the repeated measures variable(s) together with Group were modelled. This first step of the modelling process focused on analyzing potential differences among monolingual and bilingual speakers, which is the main objective of this dissertation. Subsequently, in phase 2, in order to analyze the effect of the nonverbal IQ, the receptive vocabulary, and the sentence recall abilities at the global level, three additional fixed effects were included in the best-fitting model from phase 1. If the inclusion of all, part, or any of these three predictors improved the best-fitting model of phase 1 without leading to collinearity worrying values, that model would be chosen as the best-fitting model for our data. The *variance inflation factors* (VIF) values for all the models presented in this dissertation including two or more predictors were below 2, and the tolerance values were above 0.2. Thus, our data did not reveal collinearity issues (see Field et al., 2012, p. 293 for more details).

Gender, SES and PWS subtype were not included in the models as fixed effects, since the priority was to keep the models simple. Their exclusion is justified by the fact that the analysis of the participants' characteristics did not reveal significant gender differences and that both SES and PWS subtype variables did not have much variation in our data: more than 85% of the families were middle class families and more than 73% of the participants pertained to the DEL subtype. Also, chronological age was not included as a fixed effect either given that no correlation between this factor and the standardized test raw scores was found.

Following the standard procedure within the field, we always started the phase 1 analysis with the so-called maximal model, since models "generalize best when they include the maximal random effects structure justified by the design" (Barr et al., 2013, p. 254). Thus, random intercepts for participants and for items as well as random slopes for the repeated measure(s) by participant and for Group by item were included (see Linck & Cunnings, 2015 and Winter, 2019 for more information about random effects in linguistic research). However, it is important to highlight that not all the fixed and random effects specified in the maximal model for phase 1 were kept in the best-fitting model presented for that phase, since some of them prevented the models from converging or yielded non-reliable models due to singularity issues. When the maximal model did not converge, the model was simplified until convergence was attained. There is not consensus within the field about how to proceed in this case (Linck & Cunnings, 2015), however it seems to be a standard practice to start with the simplification of the random effects structure first, although this procedure needs to be evaluated in each case individually (see Barr et al., 2013 for a complete discussion). In our analyses, following procedures in previous studies and recommendations available in statistical manuals (Barr et al., 2013; de Bruin et al., 2018; Teubner-Rhodes et al., 2016; Winter, 2019; among others), the simplification process consisted of: (i) removing the random slope interactions; (ii) decorrelating intercepts and slopes; (iii) building down the item slope structure; (iv) building down the participant slope structure. The simplification of the item slope structure was carried out first because it has been argued that the variance expected from different participants is higher than the one expected from different items (de Bruin et al., 2018, 2020; Winter, 2019). When building down both the by-item and the by-participant slope structures, one effect at a time was removed and the different possible combinations were checked for convergence. When the random-intercepts-only models did not converge or yielded singularity issues, we tried to simplify the fixed effects structure of the maximal model in a way that was meaningful to our research objectives but maintaining the maximal random structure previously detailed. Likewise, the already detailed simplification process for the random effects structure was followed with the new rebuilt model until convergence was attained. The minimal random structure accepted for conducting the analysis was the by-participant and by-item random-intercepts-only model. Once convergence was reached, while maintaining the random structure constant, the modelling process continued by simplifying the fixed effects structure to check whether the interactions (removed first) and the individual fixed effects (removed in a successive step) that were included contributed to the improvement of the model. All possible options were checked, and for each simplification, the resulting model was refitted; if any convergence or singularity problems arose, the new model was compared to the previous one using the second-order Akaike Information Criterion (AICc), which is the Akaike Information Criterion (AIC) correction for small samples (see Burnham & Anderson, 2002 for a complete review of both AIC and AICc).[34] The model with the lowest AICc value was ultimately selected as the best-fitting model, since a lower value indicates a better model fit (see Field et al., 2012, p. 263 for more details).

Model comparison is an essential part of the modelling process since as Phillips (2016, p. 208) states “a good model not only needs to fit data well, it also needs to be parsimonious. That is, a good model should be only be as complex as necessary to describe a dataset”. When the best-fitting model from phase 1 was reached, the maximal model for phase 2 was created. It included the fixed-effect(s) of the best-fitting model from phase 1 plus the three standardized measures as well as the by-participant and by-item intercepts and all the possible slopes. We included Nonverbal IQ, Receptive Vocabulary, and Sentence Recall abilities as individual fixed effects not in interaction with other predictors, since the data available did not allow us to do such analysis. Therefore, as stated in research question four, we focused on analyzing the effect of individual differences among participants at a global level, which, without any doubt, would provide us with valuable information to lay the groundwork for future research addressing a more analytical approach.

The simplification model procedure for phase 2 was the same as specified for phase 1. Starting with the maximal model, when the random-intercepts-only model did not converge or presented singularity problems, two standardized measures combinations were adopted, which resulted in three different base models (Nonverbal IQ-Receptive vocabulary, Nonverbal IQ-Sentence Recall, Receptive Vocabulary-Sentence Recall). Each model was simplified independently and the best-fitting models for each combination, when different, were compared

[34] AICc and AIC are valid to compare both nested and non-nested models (Burnham & Anderson, 2002). Even though we followed the AICc to select the best-fitting model for each model comparison, we also compared them using the AIC and both criteria coincided in the model selection process.

according to their AICc value. The best-fitting model from phase 2 was then compared to the best-fit model from phase 1; if the former had a lower AICc, then it was retained as it improved the model presented in the phase 1.

It is worth noting that due to the dummy coding system used by *R* when including categorical variables with more than two levels, the models that were built only compared the variable declared as the reference level with the rest of the variable's levels. Nevertheless, in order to have a more comprehensive understanding of how all the levels of a non-binary categorial variable compared, post hoc analysis with the Bonferroni correction were run using the *emmeans* (version 1.4.6) package.

The variance explained by the different mixed-effects models was estimated using the *r.squaredGLMM* function (MuMIn package –version 1.43.17; Barton, 2019), which specifies the marginal R^2 (variance accounted for by the fixed effects) and the conditional R^2 (variance accounted for by both fixed and random effects)[35] (see Nakagawa & Schielzeth, 2013 for a review of these measures).[36]

In the case of regression analysis, we modelled the dependent variable in a single phase, since participants only contributing one data point made the analysis simpler. In general, we included Group together with the standardized measures as predictors (multiple regression analysis). Next, we checked whether removing predictors, one at a time, improved the model. If so, the modelling process continued with the improved model until, as was the case with the mixed-effects models, further simplification stripped the model of convergence, produced singularity issues, or resulted in a model with a higher AICc. As was the case for the mixed-effects models,

[35] As Barton (2020, p. 56) defends, "R^2 is a useful goodness-of-fit measure as it has the interpretation of the proportion of the variance 'explained', but it performs poorly in model selection".
[36] As Winter (2019, p. 77) points out "R^2 is actually a measure of 'effect size'".

the three continuous predictors were *z*-scored and the categorical variable Group was contrast coded. For regression analysis, the estimation of the data variance was assessed with the multiple R-squared (data variance accounted for by the model) and the adjusted R-squared (data variance accounted for by the model taking into account the number of variables included in it) (see Field et al., 2012). Both measures were provided through the *summary* function after running the *lm* models.

Besides regression and mixed-effects models analyses, this dissertation also includes analyses of variance (repeated measures ANOVA) to address the additional objective of investigating potential differences among bilingual speakers between Spanish and Catalan results (metalinguistic and narrative tasks). ANOVA analyses were conducted combining both SPSS and JASP statistical software. Post hoc pairwise comparisons for significant ANOVA results and effect sizes were run and computed as specified in section 4.3 of this chapter. For continuous variables, the mean per participant per condition was analyzed. In the case of non-continuous dependent variables, such as the results obtained in both metalinguistic tasks (SJT and WLJT), following what seems to be a common practice within the field, the mean accuracy score by participant and condition was calculated, transforming, somehow, a binary variable into a continuous type one (see Bialystok, 1986b; Bialystok et al., 2014; Bialystok & Mitterer, 1987, among others, for a similar approach).

In this chapter we have presented a general overview of the study detailing the research questions, the standardized tests administered, the participants' cognitive and linguistic characteristics as well as the experimental tasks administered, and the statistical analyses run. The next three chapters delve into the effects of bilingualism on the EC (Chapter 5), metalinguistic (Chapter 6), and narrative abilities (Chapter 7).

5. The executive control abilities of monolingual and bilingual speakers with Prader-Willi Syndrome

The main objective of this chapter is to shed light on how monolingual and bilingual speakers with PWS compare regarding the EC abilities (research question 1). In addition, and taking an exploratory approach, we are also interested in elucidating how individual differences in terms of nonverbal IQ, receptive vocabulary, and sentence recall abilities may shape the results obtained (research question 4). To achieve these goals, we examine the participants' performance in an EC experiment (experiment 1) including two tasks: the Flanker task, and the Stroop task. As stated in Chapter 2, these two well-known and long-established experiments have been extensively used within the field to evaluate monolingual and bilingual speakers' interference suppression abilities and a relative BA has been proposed. Nonetheless, as already discussed, results are not clear-cut, especially for the young adult population (see Bialystok, 2016).

Our working hypothesis is that PWS individuals will not perform worse than their monolingual peers. In fact, if the so-called BAH holds for individuals with genetic syndromes in a comparative way to TD individuals, a positive interaction between bilingualism and EC abilities, measured through response latency (RT) and response accuracy, is expected for individuals with PWS, specially at a global level (BEPA hypothesis). Also, if it is the case that participants' individual differences for the three standardized measures included in this study play a role in the overall results, it is expected to be in terms of the nonverbal IQ and/or sentence recall (as an indirect measure of working memory) abilities. An effect of the receptive vocabulary predictor, if any, would only be expected in the Stroop task, which is the language driven experiment.

This chapter is organized as follows; we first present the participants' sample (section 5.1) followed by the experimental design of both tasks (section 5.2) and the experiment's administration procedure (section 5.3). We then describe the data analysis for the Flanker and Stroop tasks (section 5.4) as well as the results (section 5.5). The chapter concludes with a general discussion of the main findings (section 5.6).

5.1. Participants' sample

The 15 participants (eight monolinguals and seven bilinguals) described in section 4.3 completed experiment 1. Nevertheless, while the whole sample was kept for the Flanker task data analysis, as we have mentioned before, one bilingual participant was excluded from the Stroop task data analysis due to colour blindness.

Table 7. Nonverbal IQ, receptive vocabulary and sentence recall measures for monolingual and bilingual speakers in Spanish included in the data analysis of the Stroop task. Standard scores (M=100, SD 15).

	Nonverbal IQ		Receptive Vocabulary		Sentence Recall	
	Monolinguals	**Bilinguals**	**Monolinguals**	**Bilinguals**	**Monolinguals**	**Bilinguals**
Mean	75.75	92.00	64.13	70.00	66.88	68.33
SD	10.35	10.79	16.95	14.62	14.87	19.66
Range	55.00-87.00	83.00-112.00	55.00-94.00	55.00-91.00	55.00-100.00	55.00-105.00

In order to confirm that excluding one participant from the Stroop task data analysis did not compromise the group comparability (see Table 7), a mixed-design ANOVA with Standardized test (Nonverbal IQ, Receptive Vocabulary, and Sentence Recall) as the within-subjects variable and Group (Monolinguals vs. Bilinguals) as the between-subjects variable was run. Results revealed a main effect of Standardized test ($F(2, 24) = 6.13$, $p = .007$, $\eta_p^2 = 0.34$) but not of Group ($F(1, 12) = 2.55$, $p = 0.136$, $\eta_p^2 = 0.18$). The interaction between both factors was not significant either ($F(2, 24) = 0.97$, $p = .394$, $\eta_p^2 = 0.08$). Post hoc pairwise comparisons with the Bonferroni correction yielded a significant mean difference between nonverbal IQ and receptive vocabulary abilities ($p = .030$, $d = 0.80$). No differences between groups were found in terms of chronological age ($t(12) = 0.751$, $p = .467$, $d = 0.41$).

These results confirm that excluding one bilingual participant for the Stroop task data analysis did not compromise the comparability of both groups and that these results are in line with the findings presented in section 4.3 for the whole sample.

5.2. Experimental design: Flanker task and Stroop task

Experiment 1 was programmed using E-Prime software Version 2.0 (W. Schneider et al., 2002a, 2002b) and was administered using a PC laptop computer (15-inch screen; Windows 7). Participants provided their answers using a portable numpad connected to the laptop via USB. Both RT and accuracy data were collected.

Flanker task

Participants were presented with different chevron sequences and were asked to determine the direction of the central chevron. The task included three experimental conditions: (1) **Neutral** (a central chevron between two diamonds on each side; ◊◊>◊◊ or ◊◊<◊◊; (2) **Congruent** (central chevron pointing to the same direction of the other four chevrons; < < < < < or > > > > >); (3) **Incongruent** (central chevron pointing to the opposite direction of the other four chevrons; < < > < < or > > < > >). The task included a total of 96 trials (32 stimuli per condition: in 16 stimuli the central chevron pointed left and in the remaining 16 it pointed right). The 96 trials were evenly divided in two blocks of 48 trials each one (16 stimuli per condition evenly distributed for the central chevron pointing right or left) and two lists were created in order to neutralize a possible presentation bias. List A contained block 1 and block 2 (stimuli were presented with order 1-48 in each block) and list B included block 3 and block 4 (exactly the same stimuli of block 1 and 2, respectively, but in the reverse order, i.e., 48-1 in both cases). Both lists were counterbalanced across participants. Trials were presented in sequential order within each block because they were previously pseudo-randomized[37] ensuring that: (i) no two exact stimuli were presented consecutively; (ii) no more than two trials of the same experimental condition were presented in a row, and (iii) no more than two left or right answers were presented in a sequence.

Stroop task

Participants were presented with different colored words and were asked to determine the ink font color. The task included three experimental conditions: (i) **Neutral** (colored non-color words. E.g., *Ventana* 'window' written with blue ink); (ii) **Congruent** (color words in which the color word and the color font coincide. E.g., *Azul* 'blue' written with blue ink); (iii) **Incongruent** (color words in which the color word and the color font do not coincide. E.g., *Azul* 'blue' written with red ink). Three color nouns (*rojo* 'red', *azul* 'blue' and *amarillo* 'yellow') and three high frequency non-color nouns (*mesa* 'table', *hoja* 'leave' and *ventana* 'window') were used as stimuli. All bilingual participants with the exception of one participant who recognized Spanish as his/her dominant language completed the Stroop task in Catalan with the same translated stimuli. Previous research has shown that multilingual speakers show higher RTs and tend to be less accurate when completing the Stroop task in a language other than their L1 (Marian et al., 2013). Hence, in an attempt to obtain comparable measures in both groups, the stimuli were always presented in the

[37] Pseudo-randomization is not uncommon within the field and other researchers have applied similar procedures (see Costa et al., 2008 for an example on the Stroop task).

participants' dominant language. The six words included in the experiment were not cognates in Spanish and Catalan.

Participants were administered a total of 108 trials (36 trials per condition using the red, blue, or yellow ink 12 times within each one). The 108 trials were evenly divided in two blocks of 54 trials each one and, as in the previous task, two lists with inverse trial presentation order within each block were created. Participants were evenly distributed across lists. As with the Flanker task, trials were presented in sequential order within blocks because they had been previously pseudo-randomized ensuring that: (i) the same ink color was not presented twice consecutively; (ii) the same stimuli was not presented more than twice in a row, and (iii) the same experimental condition was not presented more than twice in a sequence. Both the Spanish and the Catalan versions of the task were identical with the exception of the language in which the words were presented, i.e., blocks and trial presentation order were exactly the same.

5.3. Procedure

Due to the nature of both tasks and the population targeted, we created a single EC experiment alternating one block from each task. All participants were asked to complete first a Flanker task block followed by a Stroop task block and this was repeated twice. The rationale behind this procedure was to avoid a tiredness and/or boredom effect, especially on the response latency data. Participants took short breaks between blocks.

Before completing the experimental trials, participants were presented with the general instructions of both tasks. After that, the specific instructions for the Flanker task were detailed and all doubts that the participants might have had were clarified by the author of this dissertation. They were also presented with 12 practice trials of the same nature as the Flanker task experimental trials with corrective feedback via emoticons. Once all details were clear and participants were ready to start, the first block of the Flanker task was presented. No feedback was provided for the experimental trials. Once finished, participants were provided with the Stroop task details and all possible doubts were addressed. Then, they completed the 12 practice items with corrective feedback and when they felt ready to start the first block of the Stroop task, it was administered. Before completing the second block of each task participants were presented with six practice trials. The practice trials included in the first block served to introduce the task and to make sure that the participants understood what they were expected to do. The practice trials of the second block were included to remind participants the task dynamics in order to avoid lower RTs in the first trials of the second block due to a lack of practice. This was also important to redirect attention, since as we discussed in chapter 3, some individuals may show attention deficits.

As shown in Figure 1, each trial for the Flanker task started with an *¿Estás listo/a?* 'Are you ready?' screen.[38] When participants felt ready, they pressed the space bar of the numpad and a central fixation cross, which remained on screen for 500 ms, appeared followed by the target, which was kept on screen until the participant responded.[39] Target stimuli were presented centered as images with a size of 70% by 35%. Participants were instructed to respond as quickly as possible by pressing either button 4 (chevron pointing left) or button 6 (chevron pointing right) of the numpad. The answer buttons were labelled to help participants remember the expected button answer for each case.

[38] As both tasks were administered together in a single executive control experiment, participants that completed the Stroop task in Catalan saw the *Are you ready?* screen in Catalan.

[39] While piloting the task with automatic trial presentation we detected that some of the answers provided were not intended for the trial on screen but for the previous one. In order to avoid this, we decided to keep the target on screen until the participant responded to make sure that the response entered corresponded to such trial.

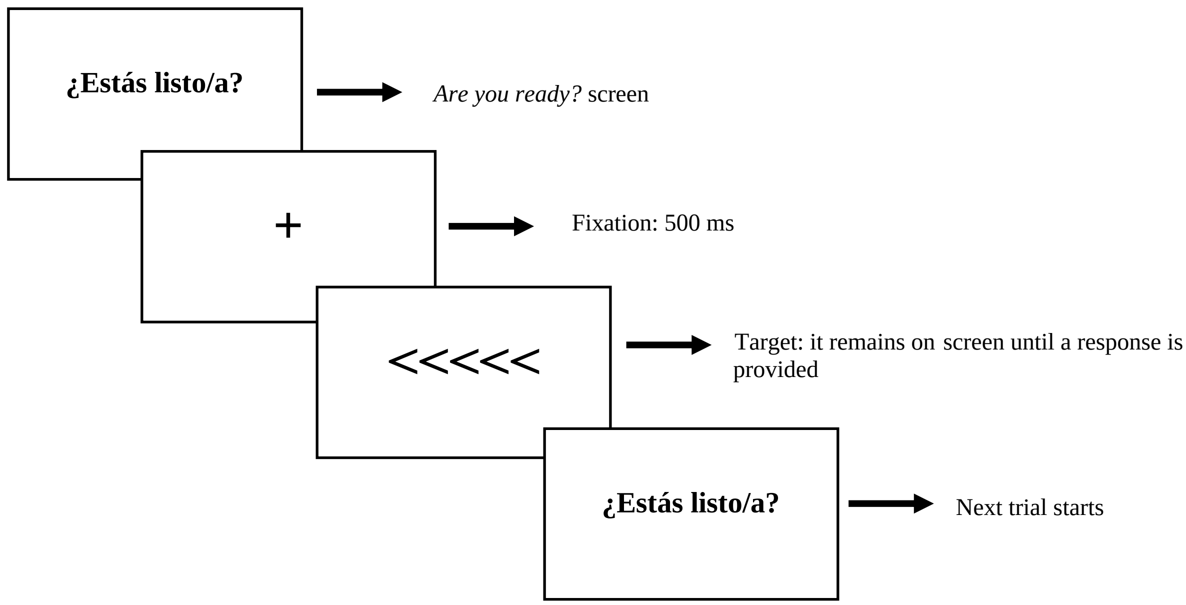

Figure 1. Trial structure for the Flanker task. A congruent stimulus is shown as an example.

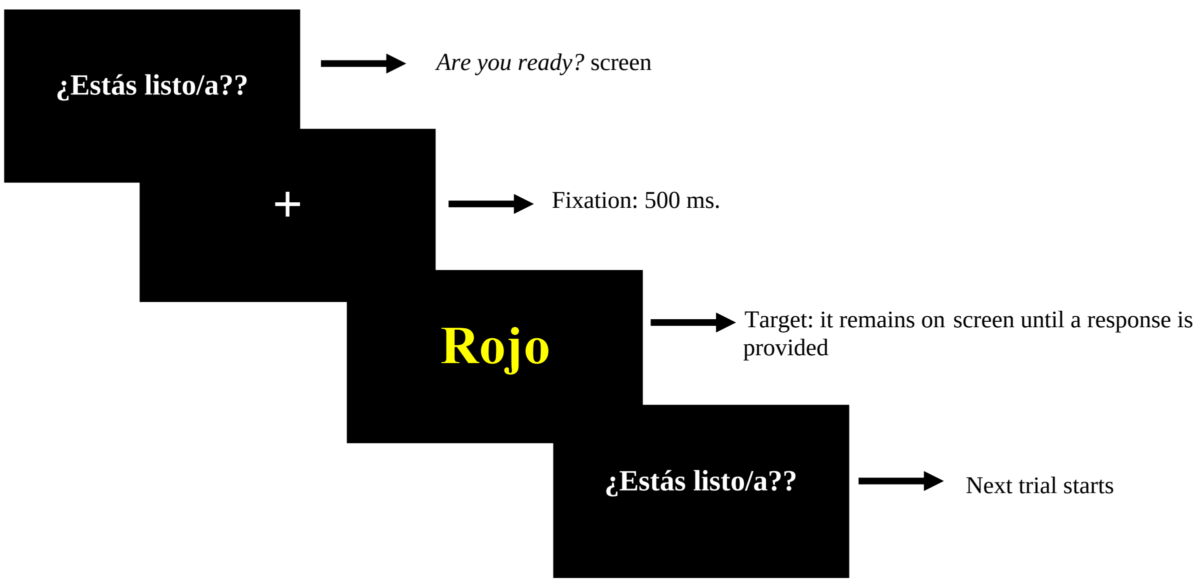

Figure 2. Trial structure for the Stroop task. An incongruent stimulus is shown as an example.

As shown in Figure 2, similarly to the Flanker task trials, each trial for the Stroop task started with an *¿Estás listo/a?* 'Are you ready?' screen followed by a central fixation cross (500 ms) and the target, which was kept on screen until a response was provided. Target stimuli were presented centered on a black screen to ensure a good color visualization in size 60 Times New Roman. Participants were instructed to respond as fast as possible by pressing either button 4 (red), button 5 (yellow) or button 6 (blue) in the numpad. Once again, the answer buttons were labelled to help participants remember the expected button answer for each ink color.

5.4. Data analysis

Both RT and accuracy data for each task were analyzed separately including all data points using a linear mixed model analysis (*lmer* model) and a generalized linear mixed model analysis (*glmer* model) respectively. Following standard practices within the field, RT data was submitted first to a cleaning process where all incorrect trials were removed and all RTs below 200 ms or above 5000 ms were excluded.[40] Finally, all mean participants' responses per condition above or below 2.5 SD were also discarded. In the case of the Flanker task a 9.65% of the data was discarded (5.76% from monolinguals and 3.89% from bilinguals) and in the case of the Stroop task a 5.55% of the data was excluded (3.30% from monolinguals and 2.25% from bilinguals). The RT dependent variable was transformed to inverse RT (-1000/RT) to reduce positive skew, which is typical in RT data (see Brysbaert & Stevens, 2018 for more details about this phenomenon and the transformation applied). During the modelling process of *lmer* models, model comparisons were performed using Maximum Likelihood. Nevertheless, the model parameters presented for the best-

[40] Researchers investigating the executive control abilities among TD individuals with tasks similar to the Flanker task and the Stroop task normally discard trials with RT above 2000 ms when working with both children (e.g., Calvo & Bialystok, 2014) and adults (e.g., Costa et al., 2009). Nonetheless, previous research analyzing the EC abilities of individuals with ID (DS population) has shown overall higher RTs. For instance, Traverso et al. (2018) administered the Flanker task to 32 individuals with DS (mean age of 14;4 years old) and results revealed a mean RT of 3230.11 ms (537.69-13822.30 ms range) for incongruent trials.

fitting model(s) were estimated with Restricted Maximum Likelihood (see de Bruin et al., 2018; Jevtovi et al., 2019 for a similar approach).

Prior to starting the actual analyses, we checked whether there was a significant effect of Block. Results showed that Block played a role on the RT results but not on the response accuracy data in both tasks. Thus, Block was included as a fixed effect when modelling RT data but not when modelling accuracy data.[41] For RT results, we first modelled the Inverse RT dependent variable with Block (Block 1, Block 2, Block 3 and Block 4) and the interaction between Experimental Condition (Neutral, Congruent, Incongruent) and Group (-0.5 for Monolinguals vs. 0.5 for Bilinguals). The maximal random effects structure included intercepts for participants and for items and a slope for the Experimental Condition by participants and slopes for Group and Block by items. The maximal model for the response accuracy data was the same as for the RT data but excluding the fixed effect of Block and its slope by item. When the best-fitting model of phase 1 was achieved, this was refitted including Nonverbal IQ, Receptive Vocabulary, and Sentence Recall as fixed effects (*z*-scored). By-participant and by-item intercepts as well as all the possible slopes were included in the random effects structure. The best-fitting model of phase 1 is presented in the results section for both RT and accuracy data (see section 4.6 for a detailed description of the modelling process followed in this dissertation). The best-fitting model of phase

[41] RT data analysis Flanker task: Reference variable: block 1. Intercept (estimate = -0.77, SE = 0.13, t = -6.04, p < .001); Block 1 vs. Block 2 (estimate = -0.10, SE = 0.03, t = -3.80, p < .001), Block 1 vs. Block 3 (estimate = -0.29, SE = 0.17, t = -1.72, p = 0.108) and Block 1 vs. Block 4 (estimate = -0.52, SE = 0.17, t = -3.05, p = .009).
RT data analysis Stroop task: Reference variable: block 1. Intercept (estimate = -0.97, SE = 0.09, t = -10.49, p < .001); Block 1 vs. Block 2 (estimate = -0.11, SE = 0.02, t = -5.05, p < .001), Block 1 vs. Block 3 (estimate = -0.26, SE = 0.13, t = -1.96, p = .072) and Block 1 vs. Block 4 (estimate = -0.36, SE = 0.13, t = -2.74, p = .018).
Accuracy data analysis Flanker task: Reference variable: block 1. Intercept (estimate = 3.86, SE = 0.61, z = 6.32, p < .001); Block 1 vs. Block 2 (estimate = 0.20, SE = 0.42, z = 0.46, p = .643), Block 1 vs. Block 3 (estimate = -0.52, SE = 0.74, z = -0.70, p = .484) and Block 1 vs. Block 4 (estimate = -0.09, SE = 0.75, z = -0.12, p = .907).
Accuracy data analysis Stroop task: Reference variable: block 1. Intercept (estimate = 5.87, SE = 0.90, z = 6.54, p < .001); Block 1 vs. Block 2 (estimate = -0.41, SE = 0.92, z = -0.45, p = .654), Block 1 vs. Block 3 (estimate = -1.31, SE = 0.91, z = -1.43, p = .152) and Block 1 vs. Block 4 (estimate = -1.31, SE = 0.91, z = -1.43, p = .152).

2 is only presented if conducted[42] and, if so, provided it is a better fit than the phase 1 model. The neutral level of the Experimental condition variable was entered into the models as the reference level with which the other two levels would be contrasted. This choice was made in order to obtain a direct comparison between neutral and incongruent trials (interference effect) and between neutral and congruent trials (facilitation effect).[43] The reference level for the fixed effect of Block was block 1. Post hoc analyses using the Bonferroni correction were run to explore the pairwise comparisons of interest not shown in the models.

5.5. Results

We first present the results of the Flanker task followed by the results of the Stroop task. In both cases RT data is presented first followed by accuracy data. For each task, descriptive data showing the group mean per condition (without previous by-participant averaging) is first discussed and inferential statistical analyses are detailed next. Descriptive RT data presents non-transformed RT values in ms while the models were run on the inverse RT. Similarly, descriptive accuracy data is presented in proportion of correct responses, but the statistical analyses were run on the binary outcome (1 = correct vs. 0 = incorrect).

---

[42] As it will be discussed in the results section, accuracy data for the Flanker task was not submitted to the phase 2 modelling process since the best-fitting model of phase 1 excluded the fixed effect of *Experimental Condition*.

[43] In this dissertation the interference effect is defined as the difference between incongruent and neutral trials and the facilitation effect as the difference between neutral and congruent trials (Coderre et al., 2013; Kousaie & Phillips, 2012a; Vinerte, 2018). Some researchers, however, have analyzed the so-called Stroop/Flanker interference effect by contrasting congruent and incongruent trials (Bialystok et al., 2008; Costa et al., 2009; among others), an approach whose appropriateness has been extensively debated due to the overlap of the interference and facilitation effects (Coderre et al., 2013; Kousaie & Phillips, 2012; Vinerte, 2018; among others).

Flanker task

Descriptive data for RT results (see Table 8) show that, overall, bilingual speakers are faster than their monolingual peers, since their RTs are lower across the board. However, despite potential global RT differences, both groups show the following speediness scale: neutral trials < congruent trials < incongruent trials. In other words, neutral trials show the lowest RTs, incongruent trials the highest RTs and congruent trials lie in between. This suggests that participants do not seem to be sensitive to a facilitation effect (RTs for congruent trials are higher than for neutral trials), but that they experience an interference effect (higher RTs for incongruent trials than for neutral trials). The interference effect is lower for bilingual speakers, which suggest they may have better interference suppression abilities. Overall, descriptive data seems to support both the BICA and the BEPA hypotheses.

Table 8. Mean RT (in ms) per condition per group (Flanker task). SD are reported in parentheses. Facilitation and interference effects are calculated taking into account the group mean per condition.

	Congruent	Neutral	Incongruent	Facilitation Effect[a]	Interference Effect[b]
Monolinguals	1427.36 (813.71)	1266.15 (594.15)	1454.47 (760.52)	-161.21	188.32
Bilinguals	1120.90 (697.08)	996.39 (471.15)	1153.54 (582.63)	-124.51	157.15

[a]A negative outcome indicates an absence of facilitation effect, since it shows that neutral trials have lower RTs than the congruent trials.
[b]A positive outcome indicates a presence of interference effect, since it shows that incongruent trials have longer RTs than the neutral trials.

In order to confirm the inferences drawn, we performed a linear mixed model analysis on the Inverse RT variable starting with the maximal model presented in the previous section. The model converged with the full fixed effect structure previously described and with the by-participant and by-item intercept. The best-fitting model for experiment 1 (see Table 9) included the fixed effects of Block and Experimental Condition and both intercepts. Group did not improve the model fit, which is interpreted as a lack of differences between monolingual and bilingual speakers. Focusing on the Experimental Condition predictor, results yielded a significant interference effect (t = 3.97, p = .028) but no facilitation effect was detected (t = 1.45, p = .246). When analyzing the effect of Block, as expected, differences were found between Block 1 vs. Block 2 (t = -3.81, p < .001) and between Block 3 vs. Block 4 (t = 8.91, p < .001) but not between Block 1 vs. Block 3 (t = -1.72, p = .109) and between Block 2 vs. Block 4 (t = 2.44, p = .177). These Block effects show, on the one hand, that participants have similar global RTs independently of the list administered (similar outcomes for blocks 1-3 and 2-4) and, on the other hand, that there is a practice effect, since participants provided faster responses when completing the second block.

The marginal R^2 of this model is 16% while the conditional R^2 is 57%. In other words, the fixed effects included in the model (Experimental Condition and Block) account for 16% of the data variance. However, when the by-participant and the by-item random effects are added, the explanatory power of the model increases to almost 60%.

Once the best-fitting model of phase 1 was achieved, we refitted it including the three standardized measures as fixed effects as well as the by-participant and by-item intercept and all the possible slopes.[44] The model converged with all the fixed effects specified (fixed effects from the best-fitting of the phase 1 plus the three standardized measures) and with the random-intercepts-only model. The best-fitting model of phase 2 included the Nonverbal IQ predictor together with Block and Experimental Condition and the by-participant and by-item random intercepts. We then compared the best-fitting model of each phase and the AICc value of phase 1

[44] As detailed in chapter 4 this constitutes the phase 2 of the modelling process and it is common for all the statistical analyses included in this dissertation including a repeated-measures design.

model was lower, which means that including the Nonverbal IQ predictor, or any of the two other standardized measures, did not improve the previous model fit.

Table 9. Model parameters for the best-fitting model of phase 1 (Flanker task). Dependent variable: Inverse RT.[45]

					Random effects	
	Fixed effects				Participant	Item
Parameters	Estimate	SE	t^a	Pr(>|*t*|)	*SD*	*SD*
Intercept	-0.83	0.13	-6.55	< .001*	0.32	0.03
Block 2	-0.10	0.03	-3.81	< .001*	-	-
Block 3	-0.29	0.17	-1.72	.109	-	-
Block 4	-0.52	0.17	-3.04	.009*	-	-
Exp. Condition Congruent	0.05	0.03	1.45	.246	-	-
Exp. Condition Incongruent	0.14	0.03	3.97	.028*	-	-
Post hoc pairwise comparisons[a]						
Block 2-Block 4	0.42	0.17	2.44	.177		
Block 3-Block 4	0.23	0.03	8.91	< .001*		

[a] For Post hoc pairwise comparisons *t* is *t-ratio*.
Note. Model reference values for Block: block 1; for Experimental Condition: neutral.
Model formula: Best_fit_phase1 <- lmer(InvRT ~ Block+ExperimentalCondition+(1|Participant) + (1|Item), data = Flanker_Task, REML = TRUE, control=lmerControl(optimizer="nloptwrap2"))

When we focus on the accuracy response results, descriptive data presented in Table 10 shows a very high accuracy rate for both monolingual and bilingual speakers in all conditions. However, the correctness of incongruent trials among monolingual speakers, albeit very high,

[45] Tabulation of all mixed-effects model results follows Linck and Cunnings's (2015) general example.

appears to be slightly lower than for bilinguals. Many studies analyzing both RT and accuracy data do not proceed with inferential statistical analyses for response accuracy when results are so close to ceiling (Grundy et al., 2017; Vinerte, 2018; among others); however, we decided to further explore the results obtained using a generalized linear mixed model analysis (see Alves-Soares, 2020; Carson, 2018 for a similar approach).

Table 10. Mean accuracy rate (proportion correct) per condition per group. Flanker task.

	Congruent		Neutral		Incongruent	
	Monolinguals	**Bilinguals**	**Monolinguals**	**Bilinguals**	**Monolinguals**	**Bilinguals**
Mean	0.97	0.98	0.97	0.96	0.90	0.95
SD	0.17	0.15	0.16	0.19	0.30	0.23

A generalized linear mixed model analysis was run on the binary Accuracy outcome with the maximal model presented in the previous section. Keeping both the by-participant and by-item intercepts, the model only converged with the fixed effect of Group.[46] The results, however, did not reveal differences between groups, which indicates that similar accuracy rates were found for monolingual and bilingual speakers (estimate = 0.11, *SE* = 0.71, z = 0.15, *p* = .877).[47] In this model, the fixed effect of Group accounts for less than 1% of the data variance and when the by-participant and by-item intercept is included, the explanatory power rises to 35%. It is important to note that while the marginal R2 is particularly low, this is not surprising if we take into account that only one fixed effect was included in the model and that the accuracy rate for both groups was very high (≥ 90%). We did not go further with this analysis, since the best-fitting model just presented

[46] Best_fitmodel_phase1 <- glmer(Accuracy ~ Group + (1|Participant) + (1|Item), data = Flanker_Task, family=binomial, control=glmerControl(optimizer="nloptwrap2")).
Group was contrast-coded (-0.5 for monolinguals and 0.5 for bilinguals).

[47] Fixed effects: intercept (estimate = 3.74, *SE* = 0.44, *z* = 8.58, p < .001).
Random effects: participant intercept (*SD* = 1.20), item intercept (*SD* = 0.54).

did not include the fixed effect of Experimental Condition. This experiment is a repeated measures design and as Winter (2011, p. 2140) states, "a repeated measures design with multiple repetitions needs at least two random effects for subjects and items and a fixed effect for repetitions". To confirm that similar between-groups results were obtained when dependencies were neutralized (i.e., only one data point per participant), a simple regression analysis with the by-participant mean accuracy (value between 0-1) as outcome variable and Group as predictor was run. Results revealed similar outcomes to the ones the mixed-effects models yielded, since the model does not significantly predict accuracy ($F(1, 13) = 0.29$, $p = .598$) and explains less than 1% of the data variance.[48]

In summary, inferential statistics for accuracy data did not reveal much, since the best-fitting model for response correctness did not include the Experimental Condition but only the fixed effect of Group and this was not significant. RT results, on the other hand, allowed us to have a better understanding of the participants' EC abilities. Results revealed a practice effect and similar outcomes for both monolingual and bilingual participants by being sensitive to an interference effect but not to a facilitation effect.

Stroop task

As shown in Table 11, descriptive data for RT suggest faster RTs in all conditions for bilingual speakers as well as smaller interference and facilitation effects for dual speakers. These results, if confirmed by inferential statistics, would show better conflict solving abilities (BICA

[48] Intercept (estimate = 0.95, $SE = 0.15$, $t = 64.09$, $p < .001$); Group (estimate = 0.02, $SE = 0.03$, $t = 0.54$, $p = .598$). Group was contrast-coded (-0.5 for monolinguals and 0.5 for bilinguals).

hypothesis) and better global executive processing abilities (BEPA hypothesis) for bilingual speakers.

Table 11. Mean RT (in ms) per condition per group (Stroop task). SD are reported in parentheses. Facilitation and interference effects are calculated taking into account the group mean per condition.

	Congruent	Neutral	Incongruent	Facilitation Effect[a]	Interference Effect[b]
Monolinguals	985.76 (333.14)	995.97 (312.63)	1140.96 (462.28)	10.21	144.99
Bilinguals	848.59 (277.56)	848.46 (285.39)	927.66 (349.97)	-0.13	79.20

[a]A negative outcome indicates an absence of facilitation effect, since it shows that neutral trials have lower RTs than the congruent trials.
[b]A positive outcome indicates a presence of interference effect, since it shows that incongruent trials have longer RTs than the neutral trials.

A linear mixed model analysis with the maximal structure detailed in the previous section was run with Inverse RT as an outcome variable. The model converged with the maximal fixed-effects structure and with the by-participants and by-items intercept. The best-fitting model (see Table 12) included the fixed effects of Block, Experimental Condition and Group as main effects and the by-participant and by-item intercepts. Results did not show a main effect of Group (t = -1.54, p = .151), which indicates that both monolinguals and bilinguals performed similarly. Results for the Experimental Condition predictor revealed an interference effect (t = 3.44, p = .002) but an absence of a facilitator one (t = -0.53, p = .601) and the effect of Block, as expected, yielded significant differences between Block 1 and Block 2 (t = -5.08, p < .001) and between Block 3 and Block 4 (t = 4.59, p < .001). Conversely, no significant differences were attested for Block 1 vs. Block 3 (t = -2.08, p = .061) and for Block 2 vs. Block 4 (t = 2.00, p = .423). Once again, participants evidenced a practice effect (higher RTs in the second block they performed) and comparable RTs independently of the list administered (similar overall RTs for blocks 1 and 3 and blocks 2 and 4).

Table 12. Model parameters for the best-fitting model of phase 1 (Stroop task). Dependent variable: Inverse RT.

					Random effects	
	Fixed effects				Participant	Item
Parameters	Estimate	SE	*t*	Pr(>|*t*|)	*SD*	*SD*
Intercept	-1.01	0.09	-11.36	< .001*	0.23	0.03
Block 2	-0.11	0.02	-5.08	< .001*	-	-
Block 3	-0.26	0.12	-2.08	.061	-	-
Block 4	-0.36	0.12	-2.90	.014*	-	-
Exp. Condition Congruent	-0.01	0.02	-0.53	.601	-	-
Exp. Condition Incongruent	0.08	0.02	3.44	.002*	-	-
Group	-0.19	0.12	-1.54	.151		
Post hoc pairwise comparisons[a]						
Block 2-Block 4	0.25	0.12	2.00	.423		
Block 3-Block 4	0.10	0.02	4.59	< .001*		

[a] For Post hoc pairwise comparisons *t* is *t-ratio*.
Note. Model reference values for Block: block 1; for Experimental Condition: neutral. Group was contrast-coded (-0.5 for monolinguals and 0.5 for bilinguals).
Model formula: Best_fit_phase1 <- lmer(InvRT ~ Block+ExperimentalCondition+Group+(1|Participant) + (1|Item), data = Stroop_Task, REML = TRUE, control=lmerControl(optimizer="nloptwrap2"))

The marginal R^2 of this model is 18% and the conditional R^2 is 49%. Namely, the fixed effects included in the model (Experimental Condition, Block and Group) account for 18% of the data variance and when including the variance associated to participants and items the explanatory power of the model increases to almost 50%.

The phase 2 model converged with the random-intercepts-only model including the three standardized measures and the three fixed effects from phase 1 model. The best-fitting model of phase 2 included intercepts for participants and for items as well as Nonverbal IQ Block, Experimental Condition and Group as fixed effects. We then compared the best-fitting model of each phase and the AICc value of the phase 1 model was lower, which, again, shows the little predictive power of the participants' cognitive and linguistic differences on the results obtained.

When analyzing response accuracy data, descriptive data reveals, once again, that both monolingual and bilingual speakers show very high accuracy rates for all conditions (see Table 13).

Table 13. Mean accuracy rate (proportion correct) per condition per group. Stroop task.

	Congruent		**Neutral**		**Incongruent**	
	Monolinguals	**Bilinguals**	**Monolinguals**	**Bilinguals**	**Monolinguals**	**Bilinguals**
Mean	0.99	0.99	0.99	1.00	0.98	0.98
SD	0.12	0.12	0.08	0.00	0.13	0.15

A generalized linear mixed model analysis was run on the Accuracy outcome variable with the maximal model described in the previous section. The model converged with the fixed effect of Experimental Condition and Group as main effects and with the random-intercepts-only model. The best-fitting model was the one including the fixed effect of Experimental Condition and both intercepts. As shown in Table 14, results suggest that participants' were significantly less accurate with incongruent trials than with neutral trials but no differences were revealed for the comparison of congruent and neutral trials. Accuracy results support the RT data findings showing an interference effect but not a facilitation one. In this model, the fixed effect of Experimental condition accounts for 10% of the data variance. However, when including the by-participant and by-item intercept, the model describes 35% of it.

Table 14. Model parameters for the best-fit model of phase one (Stroop task). Dependent variable: Accuracy (1 = correct vs. 0 = incorrect).

					Random effects	
	Fixed effects				**Participant**	**Item**
Parameters	Estimate	SE	*z*	Pr(>|*z*|)	*SD*	*SD*
Intercept	6.14	0.87	7.03	< .001*	0.88	0.71
Exp. Condition Congruent	-1.27	0.90	-1.42	.156	-	-
Exp. Condition Incongruent	-1.71	0.84	-2.03	.042*	-	-

Note. Model reference values for Experimental Condition: neutral.
Model formula:Best_fit_phase1< glmer(Accuracy~ ExperimentalCondition+(1|Participant) + (1|Item), data = Stroop_Task, family = binomial, control=glmerControl(optimizer="nloptwrap2"))

As in the previous analyses, the best-fit model of phase 1 was refitted for phase 2. The model converged with the three standardized measures and Experimental Condition as fixed effects and with the by-participant and by-item intercepts. The best-fitting model of phase 2 included the Receptive Vocabulary predictor together with Experimental Condition and the random intercepts for participants and for items. We then compared the best-fitting models of both phases and the AICc value of the phase 1 model was lower. Therefore, as it was the case with the previous analyses, the inclusion of the standardized measures did not improve the model fit of phase 1.

In summary, both RT and accuracy results revealed that both groups performed similarly in that they were sensitive to an interference effect but not to a facilitator effect.

5.6. Discussion

Experiment 1 compared the EC abilities of monolingual and bilingual individuals with PWS using both a language-based (the Stroop task) and a non-language based (the Flanker task) EC tasks. Offline data results showed a very high accuracy rate (> 90%) for both monolingual and bilingual speakers in both tasks, which reflects not only that participants understood the task and that they provided conscious responses but that their interference suppression abilities, at least measured trough response accuracy and with this experimental design, did not show signs of impairment. These results are in line with previous research showing a ceiling effect for accuracy data among TD young adults when completing both the Flanker task (Grundy et al., 2017) and the Stroop task (Vinerte, 2018).[49] The low variability attested for the accuracy data did not allow us to see much of the underlying interference and facilitation effects that the participants might experience while completing the tasks. However, the Stroop task accuracy results revealed that participants, regardless of whether they were monolingual or bilingual and despite their individual differences for three standardized measures, were sensitive to an interference effect but not to a facilitation one. This same pattern was revealed when RT data was analyzed for both experimental tasks. The fact of not detecting a facilitation effect, neither with offline nor online data, is not exceptional, since as Kalanthroff & Henik (2013, p. 413) argue it is "a much smaller and less stable effect than the interference effect". Thus, these results show that both monolingual and bilingual individuals with PWS exhibit similar interference suppression abilities in both online and offline data, and that bilingualism does not seem to affect dual speakers negatively, which is in line with previous research not finding a detrimental effect of bilingualism at the cognitive level for DS

[49] Despite the age variability of our participants, the majority of the them were adolescents or young adults (monolinguals' mean age = 22.80; bilinguals' mean age = 18.04).

individuals (Edgin et al., 2011). While no negative effects of bilingualism were detected, neither was a BA attested, which mirrors Martin's (2017) findings for and English-French bilingual adolescent with DS, who, when compared to a DS monolingual group, did not show evidence of better EC abilities. Nevertheless, it is worth noting that despite the fact that the inferential statistical analysis did not reveal significant differences between groups, descriptive data in Table 8 and Table 11 suggest a possible BA, since Spanish-Catalan speakers exhibit a smaller interference effect in both tasks, and a smaller facilitation effect in the case of the Stroop Task, two outcomes that have been argued to show positive effects of bilingualism (see Coderre et al., 2013, p. 423 for further details). This pattern of observing better bilingual descriptive results that not necessary translate into significant between-group differences has also been described in Pinto-Cardona (2017) for Spanish-English adult bilinguals with DS. We hypothesize that the differences between groups observed at the descriptive level may transfer to inferential statistics if larger samples were included. This of course is speculative, and we need to investigate whether the trend observed here holds when more and larger studies are conducted. However, the possibility that the small sample is somehow hindering a potential BA for experiment 1 cannot be completely ruled out at this time, since as Leppink et al. (2016, p. 122) state, it is common that small samples do not reveal significant within- and between-group differences even if such differences are considerable. Likewise, it could also be the case that even with more data we obtain the same results, since not finding a BA among the TD young adult population is not anecdotal but a recurrent finding (see Bialystok, 2016; Valian, 2015; Vinerte, 2018 for overviews) and previous research studying the EC abilities of these individuals using the Stroop task (Kousaie & Phillips, 2012a; Ryskin et al., 2014; among others) and the Flanker task (Grundy et al., 2017; Kousaie & Phillips, 2012b; among others) has found null effects of bilingualism. Of course, the possibility that the BAH is not as such, as suggested by different researchers (Paap et al., 2015; Paap & Greenberg, 2013), is also an option, and we will come back to this issue in chapter 8.

An effect of Block revealing that participants were faster when completing the second block of each task has also been detected. This finding suggests a practice effect, which is not surprising and is actually expected: the more experience the easier the task becomes and the quicker the RT is. Quicker responses did not lead however to higher inaccurate responses, since a preliminary analysis showed that Block was not a significant predictor of accuracy and both groups responses were at ceiling for both tasks. To examine the role of practice within the interference effect was outside the scope of this dissertation but future research should look into whether practice influences the interference suppression abilities of the PWS population similarly to the TD population. Previous research has shown that TD individuals show less interference effect in EC tasks as practice increases (see Davidson et al., 2003 for more details about practice effects for the Stroop task).

Despite having been proved that the PWS population show some EC deficiencies (Chevalère et al., 2015; Woodcock et al., 2009; among others), the effects of bilingualism found within our sample does not seem to differ much from the effects of bilingualism proposed for the TD young adult population. To scrutinize how the PWS and the TD populations might differ was out of the scope of this dissertation and it is definitely something that future research needs to address. More specifically, we believe that future studies should focus on whether and how PWS individuals differ from the TD population when facing the facilitation and interference effects on EC tasks. It could be the case that monolingual and bilingual speakers with TD and PWS show similar offline outcomes but that they differ at a processing level. Also, it could be possible that a more heterogenous sample in terms of nonverbal IQ, receptive vocabulary, and sentence recall abilities could shed light on how these three cognitive and linguistic skills may shape the EC abilities of the PWS population, if they do. Our results have not revealed a significant effect of any of these three measures. Nevertheless, it is important to highlight that despite the individual differences that may exist among participants, both groups were quite homogeneous (see section 4.3 for the complete participants' description) and this could have prevented us from seeing an actual effect. Certainly, we need more studies investigating the effects of an individual's cognitive and linguistic abilities on his/her EC skills, since previous research studying PWS monolinguals' EC abilities has revealed the important role of the nonverbal IQ abilities (Chevalère et al., 2013, 2015; among others).

6. The metalinguistic abilities of monolingual and bilingual speakers with Prader-Willi Syndrome

The main objective of this chapter is to elucidate how monolingual and bilinguals with PWS compare on different metalinguistic tasks. Additionally, we also aim at determining whether the nonverbal IQ, receptive vocabulary and sentence recall abilities play a role in the overall results (research question 4) and whether bilingual speakers perform comparably in Spanish and Catalan (research question 5). As already discussed in chapter 2, in this dissertation we adopt Bialystok's metalinguistic cognitive framework, which has been extensively used, with different language pairings and with different degrees of bilingualism, to study the linguistic and metalinguistic abilities of monolingual and bilingual children (see Friesen & Bialystok, 2012 for an overview). With the aim of contributing to this line of research with an unexplored population, we analyze the linguistic and metalinguistic abilities of our PWS participants using a Sentence Judgment Task (SJT)—experiment 2—, and a Word-Length Judgment Task (WLJT) —experiment 3—.

Given the outcomes of previous research centered on bilinguals with genetic syndromes, we do not anticipate worse outcomes for our bilingual participants (see section 2.3). If it is the case that a similar bilingualism effect holds for both TD and non-TD individuals, we hypothesize that: (i) bilingual speakers with PWS will outperform their monolingual peers in essentially metalinguistic measures but not in measures that mainly evaluate linguistic abilities; and (ii) bilingual speakers will exhibit between-language differences only for essentially linguistic measures. Based on Bialystok and Barac's (2012) results when assessing children's metalinguistic abilities (evaluation of morphological abilities), we expect that the nonverbal IQ and receptive vocabulary abilities will play a positive role in the participants' outcomes.

A detailed description of Experiment 2 is provided in section 6.1 and one of experiment 3 in section 6.2. Each section includes a description of the participants, the experimental design, the procedure, the data analysis and the results. In section 6.3, the last section of this chapter, we provide a succinct discussion of the main findings for both experiments.

6.1. Sentence Judgment task (SJT)

6.1.1. Participants' sample

The 15 participants (eight monolinguals and seven bilinguals) described in section 4.3 comprise the participants' sample for this experiment. Both monolingual and bilingual speakers completed the SJT in Spanish and the bilinguals completed the SJT in both Spanish and Catalan.

6.1.2. Experimental design

To investigate the linguistic and metalinguistic abilities of monolingual and bilingual individuals with PWS in Spanish, on the one hand, and the potential differences between Spanish and Catalan among bilingual speakers, on the other hand, two SJTs were created, one for each language. The SJTs were designed following Bialystok's extensively used English SJT task (also referred to as a grammatical judgment task). The aim of the task was to elicit the acceptability, in terms of grammaticality, of different sentences presented orally. Four experimental conditions were manipulated by crossing two factors (grammar and semantics): (i) grammatically correct and semantically correct (GCSC); (ii) grammatically correct with semantic violation (GCSV); (iii) grammar violation but semantically correct (GVSC); and (iv) grammar violation with semantic violation (GVSV). The task included six items per condition (24 items in total)[50] and participants' responses were considered to be accurate when judging GCSC and GCSV as correct and GVSV and SVSV as incorrect. Thus, correct/incorrect responses were balanced (12 expected correct items and 12 expected incorrect items). In our version of the task, out of the six items per condition included, three contained a linguistic target phenomenon related to morphology (gerund periphrasis, irregular verbs and reflexive pronouns) and the other three focused on structural order (possessor-possession, auxiliary-lexical verb and the dative-accusative clitic clusters). All sentences had between 10-13 words and only included one of the linguistic phenomena previously specified.[51] This means that if a sentence included a gerund periphrasis, no irregular verbs[52] or reflexive pronouns were included nor any possessor-possession, auxiliary-lexical verb or dative-accusative clitic clusters structures. Items were randomized and two lists were created with inverse order to avoid a potential order presentation effect. Participants were randomly and equally assigned to either one of the two versions of the task. The Catalan version included the same characteristics as the Spanish version but with different sentences.

Table 15. Summary of experimental conditions as for Bialystok's metalinguistic cognitive framework for the SJT.

Condition	Grammar	Semantics	Measure
GCSC	✓	✓	Reference
GCSV	✓	✗	Metalinguistic
GVSC	✗	✓	Linguistic
GVSV	✗	✗	---

[50] It was very important to keep the task short in order to maintain participants' attention to ensure data reliability. Thus, we followed previous research that administered this task to children including six items per condition (Bialystok, 1986b, 1992; Bialystok & Mitterer, 1987).

[51] The main aim was to keep the sentences as short as possible while at the same time ensuring adequate context to process them. Despite Spanish and Catalan being two languages that are "less condensed" than English, our two versions of the tasks were comparable to the one presented in Bialystok (1986b), where sentences include 13 words on average.

[52] Within this task, a verb is considered irregular when it is so in the tense used in the sentence. As for the verb *ser* 'to be', given its high frequency, it is not considered irregular for the purposes of this task.

Bialystok and collaborators maintain that the four aforementioned experimental conditions involve different degrees of linguistic and EC abilities as summarized in Table 15 (Bialystok, 1986b; Bialystok & Mitterer, 1987; Friesen & Bialystok, 2012; among others). Within Bialystok's metalinguistic framework (see Bialystok, 1986b; Friesen & Bialystok, 2012 for overviews), GCSC sentences are considered to be a somehow reference measure, since they are neither challenging from a linguistic nor from an EC point of view (Bialystok, 1986b). Therefore, as Bialystok suggests, if participants do not have an impaired linguistic knowledge and have understood the task, they should be able to judge these sentences as correct without too much difficulty (Bialystok, 1986b, p. 502). Our participants, both the monolinguals and the bilinguals with PWS (and the bilinguals in both languages) showed a minimum of 67% accuracy rate (4 out of 6 items) for this condition, which, a priori, does not indicate a compromised linguistic knowledge and allowed us to confirm that participants understood the task and that their responses were not at chance. GCSV sentences are considered to be a metalinguistic measure since the linguistic demands are kept low by not including ungrammaticalities but they are demanding in terms of EC: in order to answer them correctly participants need to override semantics and to focus on form and this process has been argued to be hard because "meaning is generally processed quickly and automatically" (Bialystok, 1986b, p. 502). GVSC sentences, on the contrary, are described as linguistic measures, since the EC demands are low (sentences are meaningful) but the linguistic exigences are high by introducing ungrammaticalities (Bialystok, 1986b). Finally, GVSV do not entail much difficulty, since these sentences can be judged as incorrect using both grammatical and semantic criteria. However, as Bialystok (1986b) points out, these sentences are somehow more challenging than GCSC since they imply detecting an ungrammaticality. In fact, in our interpretation of Bialystok's, individuals with higher EC abilities, i.e., those who override semantics and focus on form, would be less accurate judging these sentences as incorrect if their linguistic knowledge is "limited", since they would not be able to detect the ungrammaticality. On the contrary, those individuals with lower EC abilities whose judgment would be based mainly on meaning, would perform better because they would judge these sentences as incorrect without even considering grammar (Bialystok, 1986b, p. 502).

6.1.3. Procedure

Participants were orally presented with the 24 experimental items and were asked to determine the acceptability/correctness of them focusing on form regardless of its semantic meaning. They had two response options: good (correct) or bad (incorrect). In order to ensure that participants understood what they were expected to do, and before the administration of the task, monolingual and bilingual speakers were explained the instructions in a clear and pedagogical way, as is a common practice in the field when working with non-TD adults. As in the case of the elicitation method used by Bialystok (1986b) and Astheimer et al. (2014), participants were introduced to a fictional character (Carlitos in this case, a young man presented via a photograph). Carlitos was supposed to be recovering from an accident and, as a consequence, he sometimes said funny things that did not make sense but that were said in correct Spanish or correct Catalan. Participants were asked to help Carlitos with the evaluation of different sentences he produced and were advised to evaluate the sentences heard as correct or incorrect ignoring the meaning and focusing only on whether what Carlitos said was correct in Spanish or in Catalan. Once the instructions were clearly explained, the investigator presented the participants with examples (5) and (6).

(5) Mis padres *queremos [quieren] ir de vacaciones a la playa.
My parents $want_{1PL}$ to go of holidays to the beach.
'My parents *wants [want][53] to take a beach vacation'.

(6) No veo bien y necesito comprarme unas tijeras para ver mejor.
Not see well and need to buy to me some scissors to see better.
'I do not see well and I need to buy a pair of scissors to see better'.

When participants' did not answer *incorrect* for example *5* and *correct* for example *6*, the researcher discussed the expected answers with them and repeated the instructions in a pedagogical way: to judge the sentences according to the form and not to the meaning. Once these two examples were presented, eight practice items (two per condition) with corrective feedback were administered, as the specific objective of the training session was to make sure that the participants understood what they were expected to do. Practice items did not include the experimental target linguistic phenomena but sentences manipulating gender agreement and infinitive-clitic order as in examples 7-10.

(7) <u>Grammatically correct-semantically correct sentences (GCSC)</u>
La bandeja está encima de la nevera pero no puedo cogerla.
The tray is on top of the fridge but not can take it.
'The tray is on top of the fridge but I can't take it'.

(8) <u>Grammatically correct-semantic violation sentences (GCSV)</u>
En el museo hay muchas ballenas de pintores famosos.
In the museum there are many whales of painters famous.
'In the museum there are many whales by famous painters.'.

(9) <u>Grammar violation-semantically correct sentences (GVSC)</u>
Aquella chica *alto [alta] no irá hoy a la escuela.
That $girl_{FEM}$ $tall_{MASC}$ not will go today to the school.
'That tall[54] girl will not go to school today.'

[53] Due to the limited verbal inflexion displayed by English, the agreement error can be illustrated writing the third person singular agreement for the verb *to want*. However, in the Spanish original, the agreement error includes the first person plural inflexion (quere-**mos**) of the verb *to want*.
[54] Adjectives in Spanish agree in gender and number with the noun. In this example there is a gender violation (feminine noun and masculine adjective).

(10) Grammar violation-semantic violation (GVSV)
No puede *me recoger [recogerme] del colegio porque el elefante está en el mecánico.
Not can me pick up from the school because the elephant is in the mechanic.
'He/she cannot *me pick up [pick me up] from school because the elephant is at the mechanic'.

The investigator noted the participants' responses in an answer sheet created for that purpose. Bilingual participants were also tested in Catalan following the same protocol described for Spanish (see appendix B for a sample of the practice and the experimental items included in the Catalan version of the task). Despite the SJT being a widely used metalinguistic task within the field, our Spanish and Catalan versions of the task were piloted with five TD adults (age range 20-60 years old) in each language. After the pilot was analyzed some minor corrections were incorporated in the final tasks.

6.1.4. Data analysis

For the data analysis we carried out two different statistical analyses. The first one explored the linguistic and metalinguistic abilities of monolingual and bilingual speakers with PWS in Spanish. The second one examined the similarities and/or differences between the linguistic and metalinguistic abilities of these bilingual speakers in both Spanish and Catalan. We first present the between-group analysis (Spanish STJ analysis) followed by the within-group analysis (Spanish vs. Catalan SJT analysis).

Between-group analysis (monolingual vs. bilingual speakers)

Even though a generalized linear mixed model analysis was run including all data points, the by-group mean per condition is presented as descriptive data. We first determined whether there was an effect of List (-0.5 for List A vs. 0.5 for List B) and, as results did not reveal a significant effect, this predictor was not included in the model(s).[55] We therefore modelled the Accuracy outcome variable with Group (-0.5 for Monolinguals vs. 0.5 for Bilinguals), Experimental Condition (GCSC, GCSV, GVSC, GVSV) and Linguistic Dimension (-0.5 for Order vs. 0.5 for Morphology) and their two-way interactions as fixed effects. The maximal random effects structure included intercepts for participants and for items and a slope for Experimental Condition in interaction with Linguistic Dimension by participants and a slope for Group by items. In a second stage, we refitted the best-fitting model of phase 1 maximizing it by including Nonverbal IQ, Receptive Vocabulary, and Sentence Recall as fixed effects (*z*-scored). By-participant and by-item intercepts as well as all the possible slopes were included in the random effects structure. The best-fitting model of phase 1 is presented in the results section (see section 4.6 for a detailed description of the modelling process followed in this dissertation). The best-fitting model of phase 2 is presented next.

The GCSV level (metalinguistic measure) of the Experimental Condition variable was entered into the model as the reference level with which the other three levels would be contrasted. This choice was made in order to obtain a direct comparison between GCSV (metalinguistic measure) and GVSC (linguistic measure) so that we could shed light on how the linguistic and metalinguistic abilities of both monolinguals and bilinguals compare. Post hoc analysis using the Bonferroni correction were run to explore the rest of the pairwise comparisons.

Within-group analysis (Spanish vs. Catalan among bilingual speakers)

In order to explore potential similarities/differences between languages among bilingual speakers, the mean per participant per condition was submitted to a repeated measures ANOVA

---

[55] Intercept (estimate = 1.22, $SE = 0.32$, $z = 3.80$, $p < .001$); List (estimate = 0.61, $SE = 0.46$, $z = 1.33$, $p = .183$).

and significant effects were further explored with post hoc pairwise comparisons using the Bonferroni correction. If the sphericity condition was not met, the degrees of freedom were adjusted using either the Greenhouse-Geisser correction ($\varepsilon < .75$) or the Huynh-Feld correction ($\varepsilon > .75$).

6.1.5. Results

Descriptive data is presented first, followed by the inferential statistical analyses for both the between-group and the within-group analyses.

Between-group analysis (monolingual vs. bilingual speakers)

Descriptive data presenting the mean score per group per condition (with previous by-participant averaging) is included in Table 16.

Table 16. Accuracy rate (maximum score = 6) per condition per group (experiment 2: SJT).

	GCSC		GCSV		GVSC		GVSV	
	Monolinguals	Bilinguals	Monolinguals	Bilinguals	Monolinguals	Bilinguals	Monolinguals	Bilinguals
Mean	5.88	5.57	3.75	3.86	4.00	2.86	4.63	3.86
SD	0.35	0.79	1.98	1.77	0.76	2.34	1.30	2.48

Starting with a within-group analysis and adopting a one point difference criterion between means as a possible indicator of potential differences, we can see that both monolingual and bilingual speakers seem to show higher means –which translates into higher accuracy– when evaluating GCSC sentences than when judging any of the other three sentence types. Monolingual speakers do not seem to differ much when comparing GCSV, GVSC and GVSV. However, in the case of the bilingual speakers, it appears that they are better at ignoring semantic anomalies (GCSV-metalinguistic measure) than at detecting linguistic problems (GVSC-linguistic measure) and that they do better at rejecting ungrammatical sentences when they also contain a semantic anomaly than when the only problem is a grammar violation. In a between-group analysis, we find a one-point difference only for the GVSC condition, which seems to suggest a higher sensibility to ungrammatical sentences on the part of the monolinguals.

When examining the Linguistic dimension variable (see Table 17), a within-group analysis comparing the mean scores for order and morphology suggests that only monolingual speakers would be sensible to this manipulation and that they would be more accurate judging GVSC sentences when they include ungrammaticalities related to order than to morphology. No potential differences between groups for each linguistic dimension seem to arise. As regards differences between conditions within each linguistic dimension per group, variances of at least one point are detected for GCSC-GVSC and GCSC-GVSV in bilingual speakers when judging order and between GCSC-GCSV and GCSC-GVSC for both monolinguals and bilinguals when judging morphology.

Table 17. Accuracy rate (maximum score = 3) per condition and linguistic dimension per group (experiment 2: SJT).

	GCSC		GCSV		GVSC		GVSV	
	Monolinguals	Bilinguals	Monolinguals	Bilinguals	Monolinguals	Bilinguals	Monolinguals	Bilinguals
ODER								
Mean	2.88	2.71	2.00	2.14	2.50	1.57	2.25	1.71
SD	0.35	0.49	1.07	0.69	0.76	1.27	0.89	1.38
MORPHOLOOGY								
Mean	3.00	2.86	1.75	1.71	1.50	1.29	2.38	2.14
SD	0.00	0.38	1.16	1.25	0.53	1.25	0.74	1.21

In order to confirm whether the differences discussed above were statistically significant, a generalized linear mixed model analysis was run on the Accuracy outcome variable. The maximal model described in the previous section did not converge and convergence without singularity issues was attained with all the original fixed effects and the random-intercepts-only model. The best-fitting model for phase 1 included the fixed effects of Experimental Condition and Group (not in interaction), and the by-participant and by-item intercepts. The Linguistic Dimension predictor, alone or in interaction, did not improve the model fit, and was therefore excluded to calculate the model parameters.

As shown in Table 18, results did not reveal a main effect of Group (z = -1.16, p = .248). Independently of being monolingual or bilingual, participants were significantly more accurate judging GCSC sentences than GCSV sentences (z = 4.39, p < .001), GVSC (z = 4.81, p < .001), or GVSV (z = 3.74, p = .001). These findings not only indicate that our participants were significantly more accurate judging grammatically correct and semantically sound sentences than any other sentence types but also that they understood the task and that their responses were not by chance. Results did not reveal a different mastering of the metalinguistic and linguistic abilities among participants (z = -0.65, p = .513).

The marginal R^2 of this model is 27% while the conditional R^2 is 39%. In other words, the fixed effects included in the model (Experimental Condition and Group) account for the 27% of the data variance. However, when adding the by-participant and the by-item random effects the model accounts for almost 40% of the variance. However, 60% still remains unexplained. This is not surprising, since as Winter (2019, p. 77) points out, it is not uncommon, and actually it is somehow expected in linguistic studies to obtain relatively low R^2 values, since in this author's words "language is complex and humans are messy, so our models rarely account for that much variance".

Table 18. Model parameters for the best-fitting model of phase 1 for experiment 2 (SJT in Spanish). Dependent variable: Accuracy (0= incorrect, 1= correct).

					Random effects	
	Fixed effects				Participant	Item
Parameters	Estimate	SE	*z* [a]	Pr(>|*z*|)	*SD*	*SD*
Intercept	0.62	0.34	1.80	.071	0.69	0.42
Experimental Condition GCSC	2.73	0.62	4.39	< .001*	-	-
Experimental Condition GVSC	-0.27	0.41	-0.65	.513	-	-
Experimental Condition GVSV	0.40	0.42	0.96	.339	-	-
Group	-0.52	0.44	-1.16	.248	-	-
Post hoc pairwise comparisons						
GCSC-GVSC	3.00	0.62	4.81	< .001*		
GCSC-GVSV	2.33	0.63	3.74	.001*		
GVSC-GVSV	-0.67	0.42	-1.60	.658		

[a] For Post hoc pairwise comparisons *z* is *z-ratio*.
Note. Model reference value for Experimental Condition: GCSV. Group factor was contrast-coded (monolinguals, coded as -0.5 vs. bilinguals, coded as 0.5). Model formula: best_model_fit_phase1 <- glmer(Accuracy~ExperimentalCondition+Group+(1|Participant)+(1|Item),data=SJTSP_Test,family=binomial,control =glmerControl(optimizer="nloptwrap2")).

In order to see whether the participants' nonverbal IQ, receptive vocabulary, and sentence recall abilities might explain differences on the participants' overall accuracy rate, these three measures where entered into the best-fitting model of phase 1 as fixed effects. The random structure included the by-participant and by-item intercept as well as a random slope for condition by participant and random slopes for the three standardized measures by item. Convergence without singularity issues was not attained even with the random-intercepts-only model. Thus, three base models with two standardized measure combinations were created. The only model that converged without issues with the random-intercepts-only model was the combination including both Nonverbal IQ and Receptive Vocabulary predictors. The best-fitting model for phase 2 was the one including the Nonverbal IQ predictor as well as Experimental Condition and Group. The best-fitting models of phase 1 and phase 2 were compared and the latter was revealed as a better model fit. Therefore, this model was kept as the best-fitting model for experiment 3 and it is presented below.

As presented in Table 19, results for the phase 2 model show similar outcomes for the Experimental Condition and Group predictors to those previously discussed and reveals the Nonverbal IQ abilities as a significant Accuracy predictor ($z = -1.99$, $p = .047$), i.e., a higher nonverbal IQ leads to lower accuracy rates. These results are puzzling and not in line with previous research. For instance, Bialystok and Barac's (2012) showed a significant positive predictive effect of nonverbal IQ abilities when evaluating the metalinguistic abilities of English-Hebrew bilinguals, albeit with a different metalinguistic task (essentially a linguistic task evaluating participants' morphological knowledge) and with a different population (children).

In this case, the marginal R^2 is 30% while the conditional R^2 is 39%. In comparison to the previous model, including the nonverbal IQ abilities as a predictor increased the variance accounted for by the fixed effects by 3%. Similar R^2 values have been found in previous research using regression models to analyze the participants' metalinguistic abilities (e.g., Bialystok & Barac, 2012, p. 71).

Table 19. Model parameters for the best-fitting model of phase 2 for experiment 2 (SJT in Spanish). Dependent variable: Accuracy (0= incorrect, 1= correct).

					Random effects	
	Fixed effects				**Participant**	**Item**
Parameters	Estimate	SE	z^a	Pr(>|z|)	*SD*	*SD*
Intercept	0.63	0.33	1.94	.052	0.55	0.42
Nonverbal IQ	-0.48	0.24	-1.99	.047*	-	-
Experimental Condition GCSC	2.74	0.62	4.39	< .001*	-	-
Experimental Condition GVSC	-0.27	0.41	-0.65	.513	-	-
Experimental Condition GVSV	0.40	0.42	0.96	.339	-	-
Group	0.06	0.48	0.12	.902	-	-
Post hoc pairwise comparisons[a]						
GCSC-GVSC	3.00	0.62	4.81	< .001*		
GCSC-GVSV	2.34	0.63	3.73	.001*		
GVSC-GVSV	-0.67	0.42	-1.60	.657		

[a] For Post hoc pairwise comparisons *z* is *z-ratio*.
Note. Model reference value for Experimental Condition: GCSV. The Group predictor was contrast-coded (monolinguals, coded as -0.5 vs. bilinguals, coded as 0.5) and the Nonverbal IQ predictor was *z*-scored.
Model formula: Best_model_fit_phase2 <- glmer(Accuracy ~ ScaleNvIqstandard + ExperimentalCondition + Group + (1|Participant) + (1|Item), data=SJTSP_Test, family=binomial, control=glmerControl(optimizer="nloptwrap2"))

Within-group analysis (Spanish vs. Catalan among bilingual speakers)

As shown in Table 20, and mainly focusing in a between language analysis, descriptive data does not seem to suggest overall differences between languages. Likewise, no differences between languages are revealed when dissociating order from morphology linguistic phenomena (see Table 21)

Table 20. Accuracy rate (maximum score = 6) per condition and language among bilingual speakers (experiment 2: SJT).[56]

	SPANISH				CATALAN			
	GCSC	GCSV	GVSC	GVSV	GCSC	GCSV	GVSC	GVSV
Mean	5.57	3.86	2.86	3.86	5.57	4.57	2.43	3.43
SD	0.79	1.77	2.34	2.48	0.53	1.90	1.99	2.23

Table 21. Accuracy rate (maximum score = 3) per condition, linguistic dimension and language among bilingual speakers (experiment 2: SJT).

	GCSC		GCSV		GVSC		GVSV	
	Order	Morphology	Order	Morphology	Order	Morphology	Order	Morphology
SPANISH								
Mean	2.71	2.86	2.14	1.71	1.57	1.29	1.71	2.14
SD	0.49	0.38	0.69	1.25	1.27	1.25	1.38	1.21
CATALAN								
Mean	2.86	2.71	2.29	2.29	1.57	0.86	1.71	1.71
SD	0.38	0.49	0.95	0.95	0.98	1.21	1.25	1.11

In order to verify that bilinguals performed similarly in Spanish and Catalan, the by-participant mean score per condition was submitted to a repeated measures ANOVA with Experimental Condition (GCSC, GCSV, GVSC, GVSV), Linguistic Dimension (Order and Morphology) and Language (Spanish and Catalan) as the within-subjects variables. The Mauchly's test indicated that the sphericity assumption was not met for the Experimental Condition × Language interaction ($\chi 2(5) = 17.93$, $p = .004$), thus degrees of freedom were adjusted for this effect using the Greenhouse-Geisser correction ($\varepsilon = .42$). Results only revealed a main effect of the Experimental Condition ($F(3, 18) = 7.30$, $p = .002$, $\eta_p^2 = 0.55$) but not of Language ($F(1, 6) = 0.01$, $p = .927$, $\eta_p^2 = 0.00$) nor of the Linguistic Dimension ($F(1, 6) = 0.84$, $p = .395$, $\eta_p^2 = 0.12$).

[56] Spanish data for bilingual speakers has already been presented in tables 16 and 17. Nonetheless, it has been deemed appropriate to repeat this information in tables 20 and 21 to present both language results contrastively.

The interactions between these factors were not significant either: Language × Experimental Condition (F(1.25, 7.49) = 0.26, p = .677, η_p^2 = 0.04), Linguistic Dimension × Experimental Condition (F(3, 18) = 2.24, p = .118, η_p^2 = 0.27), Language × Linguistic Dimension (F(1, 6) = 2.27, p = .182, η_p^2 = 0.28), and Language × Linguistic Dimension × Experimental Condition (F(3, 18) = 1.05, p = .395, η_p^2 = 0.15). Post hoc pairwise comparisons with the Bonferroni correction showed a significant difference between GCSC and GVSC (p = .028, d = 1.66) and a marginally significant difference between GCSC-GVSV (p = .058, d = 1.41). It is worth noting that although the post hoc pairwise comparisons did not reach statistical significance for GCSC-GCSV (p = .106, d = 1.23), possibly due to the small sample, the effects size is large. Based on what the mixed-effects models analysis has revealed for Spanish and the lack of significance for the language variable, it seems safe to conclude that in Catalan, as was the case for Spanish, and independently of the linguistic dimension included (order vs. morphology), bilinguals show a higher accuracy rate when judging GCSC sentences (no grammatical or semantics issues) than with any other type of sentence.

In summary, the between-group analysis revealed that both monolinguals and bilinguals obtained overall comparable accuracy rates and that they were both more accurate judging correct and semantically sound sentences than any other type of sentence. Also, the nonverbal IQ abilities were revealed as a potential accuracy predictor. However, contrary to our expectations, better IQ abilities seem to result in lower accuracy rates. The within-group analysis among bilingual speakers did not reveal different outcomes for Spanish and Catalan.

6.2. Word-Length Judgment task (WLJT)

6.2.1. Participants' sample

As in experiment 3, the eight monolingual speakers and the seven bilingual speakers described in section 4.3 completed the WLJT in Spanish. Bilinguals were also administered experiment 4 in Catalan.

6.2.2. Experimental design

To complement experiment 3, a second metalinguistic task was created to have a better understanding of the participants' metalinguistic abilities. As a result, a modified version of the word-length judgment task-type presented in Bialystok (1986a), Ricciardelli (1992) and Yelland et al. (1993) was designed in Spanish and Catalan. In this task, as in Bialystok's (1986a), participants were presented with a two-word pair and were asked to identify the longest word of the pair. Four experimental conditions resulted from the referent size and word length manipulation. Each condition included six items, which made a total of 24 items (see examples 11-14):[57]

(11) Congruent pairings (largest animal/object was the longest word)

Hipopótamo-Gato or Helicóptero-Libro

'Hippopotamus-Cat' or 'Helicopter-Book'

(12) Incongruent pairings (smallest animal/object was the longest word)

Mariposa-Vaca or Teléfono-Coche

'Butterfly-Cow' or 'Telephone-Car'

[57] As it was the case with experiment 3, the examples provided are from the Spanish version of the task (see appendix C for a sample of practice and experimental items of the Catalan version of the task).

(13) Big Neutral pairings (big animals/objects independently of the word length)

Leopardo-Tigre or Congelador-Horno

'Leopard-Tiger' or 'Freezer-Oven'

(14) Small Neutral pairings (small animals/objects independently of the word length)

Mariquita-Mosca or Micrófono-Peine

'Ladybug-Fly' or 'Microphone-Comb'

Each condition included three [+ animate] pairs (real or fictitious animals) and three [-inanimate] pairs (concrete nouns: objects). Both words of the pair were either masculine or feminine. In the Spanish version, short words had two syllables and long words had between four and five syllables. Given that in Catalan monosyllables are more common than in Spanish, in the Catalan version of the task short words had one syllable and long words had between three and four syllables. In both languages there was a minimum of two syllable difference between short and long words (see Yelland et al., 1993 for a similar approach).

Besides the experimental task, participants were asked to complete a posttest task whose objective was to ensure that they were familiar with all the words included in the experiment in order not to override a potential semantics interference by lack of awareness of the meaning of a specific word. The posttest consisted of a PowerPoint presentation showing photos of all the objects and animals included in the task. Participants were presented with an image at a time and were asked to name the object/thing or the animal depicted. If participants did not know a specific word, the experimental item containing that word was excluded from the analysis. For the Spanish version of the task, only a 0.6% of the items were excluded (two bilingual speakers ignored one word each). For the Catalan version of the task, a 2.4% of the cases were excluded (only four words were ignored in total).

Similarly to Yelland et al. (1993), the size of the animals and objects was determined according to real world perception and given the possible subjective interpretation of size, following these authors' procedure, we conducted first a pilot study. In our case, two different pilot studies (one in Spanish and one in Catalan), with two phases in each one, were conducted among TD adult individuals to ensure the size (di)similarity between the two words included in each word-pair objectively. In phase one of the pilot study, different word-pairs (26 in Spanish and 24 in Catalan) were presented via an online platform to 27 Spanish speaking participants and to 27 Catalan dominant bilingual participants.[58] Participants were asked to complete a multiple choice questionnaire specifying whether they thought that the two words presented depicted objects or animals of similar size or whether one of the two words represented a bigger object or animal. All those pairs that reached a minimum agreement of 65% were included as experimental items. Nonetheless, given that, in both Spanish and Catalan pilot tests, less than six pairs per condition reached this threshold, a second pilot test was administered in both languages to complete the experimental items remaining. In this second pilot test, 20 Spanish speaking TD adults and 18 Catalan dominant TD adult speakers[59] were presented with 20 new word-pairs in Spanish and 19 word-pairs in Catalan. The procedure and item inclusive criteria for pilot test 2 were exactly the same as in the pilot test 1. For those conditions where more than six potential word-pairs finally reached the 65% inclusive threshold, the six pairs with the highest agreement percentage were included.

[58] In pilot test 1, 63% of the Spanish speaking participants were between 20-40 years old, 29.6% were between 41-60 and 7.4% were above 61. In the case of Catalan dominant bilingual speakers, 81.5% were between 20-40 years old and 18.5% between 41-60 years old.

[59] In pilot test 2, 70% of the Spanish speaking participants were between 20-40 years old, 20% were between 41-60 and 10% were above 61. In the case of Catalan dominant bilingual speakers, 83.3% were between 20-40 years old, 11.1% between 41-60 years old and 5.6% were above 61.

Items were randomized and the order of presentation of long and short words was counterbalanced in each condition. This means that in each condition and for three word-pairs the long word was presented first and for the other three the short word was said first. As in experiment 3, two lists were created with inverse order to avoid a potential order of presentation effect. Participants were randomly and evenly assigned to either one of the two lists. The Catalan version of the task included the same characteristics as the Spanish one but with different word-pairs.

6.2.3. Procedure

Following the procedure described in Yelland et al. (1993), participants were presented first with the difference between long and short words and, as these authors suggested, long words were described as words that took longer to say than short words. Different examples such as *escritorio* vs. *reloj* 'desk vs. watch' were presented to familiarize participants with the task's dynamics. After this short phase of familiarization, following the original procedure described in Yelland et al. (1993), participants were asked to tell the investigator two long words and two short words to make sure they understood the task; they received feedback and their answers were discussed. Once the investigator was confident that the participants mastered the long-short word difference, the training session with corrective feedback began with six different long-short word-pairs including verbs (e.g., *pinta-paseamos* 'draw$_{3SG}$-stroll$_{1PL}$', *bailas-clasifican* 'dance$_{2SG}$-classify$_{3PL}$). During the experimental session, participants were presented orally with the 24 word-pairs and were asked to choose the longest word of each pair. No feedback was provided during this part of the task. Once the task finished, participants were asked to complete the posttest. The investigator marked the participants' responses for both the experimental task and the posttest in an answer sheet created for that purpose.

6.2.4. Data analysis

The data analysis for experiment 4 follows the same outline described for experiment 3. We first present the between-group analysis (Spanish WLJT analysis) and next the within-group analysis comparing the results of bilingual speakers in Spanish and Catalan.

Spanish WLJT analysis (monolingual and bilingual speakers)

As in experiment 3, the Accuracy binary dependent variable (1= correct, 0= incorrect) was modelled using a generalized linear mixed model analysis. We first checked the effect of List and as it was not significant it was not included in the model(s) as a fixed effect.[60] Phase 1 of the modelling process incorporated Group (-0.5 for Monolinguals vs. 0.5 for Bilinguals), Experimental Condition (Congruent, Incongruent, Big Neutral, Small Neutral), and animacy (-0.5 for Animate vs. 0.5 for Inanimate) and their two-way interactions as fixed effects. The maximal random effects structure included intercepts for participants and for items and slopes for the interaction between Experimental Condition and Animacy for participants, and a slope for Group for items. The best-fitting model of phase 1 was maximized in phase 2 by including the three standardized measures (z-scored) as fixed effects as well as the by-participant and by-item intercepts and all the slopes possible. Congruent items were entered into the model as the reference level for the Experimental Condition variable. Post hoc analysis using the Bonferroni correction were run to explore the rest of the pairwise comparisons.

[60] Intercept (estimate = 5.99, SE = 1.65, z = 3.63, p < .001); List (estimate = -1.46, SE = 1.73, z = -0.84, p = .401).

Spanish vs. Catalan WLJT analysis (bilingual speakers)

Following the same procedure as in the previous experiment, data analysis to compare bilinguals' results in Spanish and Catalan was conducted on the mean per participant per condition using a repeated measures ANOVA. All significant main effects were further analyzed with pairwise comparisons using the Bonferroni correction.

6.2.5. Results

Descriptive data is presented first followed by the inferential statistics for both the between-group and the within-group analyses.

Between-group analysis (monolingual vs. bilingual speakers)

As in experiment 3, all data points were included in the analysis and the by-group mean score per condition (with previous by-participant averaging) is presented in the tables below. As displayed in Table 22, both monolinguals and bilinguals show a very high accuracy rate for all the conditions. Likewise, no sign of a possible role of animacy is detected, since, as shown in Table 23, both groups maintain a high level of accuracy when judging the longest word of the pair regardless of the [± animate] feature.

Table 22. Accuracy rate (in percentages)[61] per condition per group (experiment 3: WLJT).

	Congruent		**Incongruent**		**Big Neutral**		**Small Neutral**	
	Monolinguals	**Bilinguals**	**Monolinguals**	**Bilinguals**	**Monolinguals**	**Bilinguals**	**Monolinguals**	**Bilinguals**
Mean	100	97.62	91.67	97.62	91.67	100	100	97.62
SD	0.00	6.30	23.57	6.30	17.82	0.00	0.00	6.30

[61] Scores are presented in percentages in tables 22-25 since one item in two bilingual participants' answers had to be excluded from the analysis.

Table 23. Accuracy rate (in percentages) per condition and animacy per group (experiment 3: WLJT).

	Congruent		Incongruent		Big Neutral		Small Neutral	
	Monolinguals	Bilinguals	Monolinguals	Bilinguals	Monolinguals	Bilinguals	Monolinguals	Bilinguals
ANIMATE								
Mean	100	95.24	95.83	100	87.50	100	100	95.24
SD	0.00	12.60	11.78	0.00	24.80	0.00	0.00	12.60
INANIMATE								
Mean	100	100	87.50	95.24	95.83	100	100	100
SD	0.00	0.00	35.36	12.60	11.78	0.00	0.00	0.00

Starting with the maximal model described in the previous section, the Accuracy binary outcome variable was modelled using a generalized mixed model analysis. The model did not converge even with the random-intercepts-only model. As a result, the fixed-effects structure was simplified by removing the interactions including Animacy. The maximal random-effects structure was also modified accordingly by breaking down the interaction of the by-participant random slope. Again, the random-intercepts-only model did not converge. In an attempt to obtain convergence, a new base model was created including Group, Experimental Condition and Animacy as fixed effects (no interactions) and maintaining the maximal random-effects structure possible. Convergence without singularity issues was attained with the random-intercepts-only model; further simplification of the fixed-effects structure was conducted keeping the random structure constant. The best-fitting model for phase 1 only included the fixed effect of Animacy, which was not significant (estimate = 0.25, *SE* = 0.83, *z* = 0.30, *p* = .766).[62]

[62] Model formula: Best_fit_model_WLJ <- glmer(Accuracy ~ Animacy + (1|Participant) + (1|Item), data = WJT_Test, family=binomial, control=glmerControl(optimizer="nloptwrap2"))
Fixed effects: intercept (estimate = 6.10, *SE* = 1.83, *z* = 3.34, *p* = .001).
Random effects: participant intercept (*SD* = 2.60), item intercept (*SD* = 0.96).

The variance explained by the predictor Animacy is less than 1%. However, when adding the by-participant and by-item intercepts it increases to 70%. Basically, these results show that the experimental variables manipulated in this experiment (Group, Experimental Condition, and Animacy) had very little explanatory power to describe the results obtained, which is not surprising if we take into account the high and similar results obtained by both groups for all conditions (see descriptive results in Table 22 and Table 23).

The best-fitting model of Phase 1 was next refitted including the three standardized measures (*z*-scored) and the by-participant and by-item intercepts as well as all the slopes possible. The model converged with the random-intercepts-only model. Further simplification of this model revealed that the best-fitting model for phase 2 included Animacy and the Sentence Recall predictors. However, a comparison between the phase 1 and the phase 2 models showed that the model which did not include the Sentence Recall predictor was a better model fit.

Within-group analysis (Spanish vs. Catalan between bilingual speakers)

Descriptive data comparing the mean score per condition in Spanish and Catalan is presented in Table 24. Participants show very high and comparable accuracy rates for all conditions in both languages and the predictor Animacy does not seem to play a role when determining the longest word of the pair (see Table 25).

Table 24. Accuracy rate (in percentages) per condition and language among bilingual speakers (experiment 3: WLJT).[63]

	Congruent		Incongruent		Big Neutral		Small Neutral	
	Spanish	**Catalan**	**Spanish**	**Catalan**	**Spanish**	**Catalan**	**Spanish**	**Catalan**
Mean	97.62	97.62	97.62	94.76	100	100	97.62	95.24
SD	6.30	6.30	6.30	9.00	0.00	0.00	6.30	12.60

Table 25. Accuracy rate (in percentages) per condition, animacy and language among bilingual speakers (experiment 3: WLJT).

	Congruent		Incongruent		Big Neutral		Small Neutral	
	Animate	**Inanimate**	**Animate**	**Inanimate**	**Animate**	**Inanimate**	**Animate**	**Inanimate**
SPANISH								
Mean	95.24	100	100	95.24	100	100	95.24	100
SD	12.60	0.00	0.00	12.60	0.00	0.00	12.60	0.00
CATALAN								
Mean	95.24	100	90.48	100	100	100	90.48	100
SD	12.60	0.00	16.26	0.00	0.00	0.00	25.20	0.00

In order to confirm the inferences presented above, the by-participant mean accuracy percentage was submitted to a repeated measures ANOVA with Experimental Condition (Congruent, Incongruent, Big Neutral, Small Neutral), Animacy (Animate and Inanimate) and Language (Spanish and Catalan) as the within-subjects variables. Results did not reveal any significant main effect: Experimental condition ($F(3, 18) = 0.60$, $p = .626$, $\eta_p^2 = 0.09$), Animacy ($F(1, 6) = 2.40$, $p = .172$, $\eta_p^2 = 0.29$), Language ($F(1, 6) = 2.40$, $p = .172$, $\eta_p^2 = 0.29$). Similarly, any of the interactions between variables was significant: Experimental Condition × Animacy ($F(3, 18) = 0.49$, $p = .695$, $\eta_p^2 = 0.08$), Experimental Condition × Language ($F(3, 18) = 0.63$, $p = .604$, $\eta_p^2 = 0.10$), Animacy × Language ($F(1, 6) = 1.78$, $p = .231$, $\eta_p^2 = 0.23$) and Experimental

[63] Spanish data for bilingual speakers has already been presented in tables 22 and 23. Nonetheless, as for experiment 3, we repeat this information in tables 24 and 25 to present the reader with a direct comparison between Spanish and Catalan results for bilingual speakers.

Condition × Animacy × Language ($F(3, 18) = 1.95$, $p = .158$, $\eta_p^2 = 0.25$). These results show similar outcomes in both Spanish and Catalan for all the conditions tested and independently of the [± animate] feature.

Overall, the results of experiment 4 show that both monolingual and bilingual speakers are able to discriminate form from meaning in a very effective way independently of the manipulation of the Experimental Condition and the Animacy feature and, in the case of bilingual speakers, this holds for both their dominant and their non-dominant language.

6.3. Discussion

In this chapter we have presented the results of two metalinguistic tasks that pose different demands on the participants' EC abilities investigated, namely the linguistic and metalinguistic abilities of the PWS population. In Experiment 2 (SJT task) participants were presented with different types of sentences and were asked to determine their grammaticality. Four conditions were manipulated according to grammar and semantics: GCSC (correct and meaningful sentences), GCSV (correct but meaningless sentences), GVSC (ungrammatical but meaningful sentences) and GVSV (ungrammatical and meaningless sentences). Two of these conditions were especially relevant in assessing the participants' linguistic (GVSC sentences) and metalinguistic (GCSV sentences) abilities and previous research has consistently shown a BA for TD bilingual children in the latter case (see Bialystok, 2001; Friesen & Bialystok, 2012, for overviews). Our results evidenced that all participants, independently of being bilingual or monolingual speakers were significantly more accurate when judging grammatically correct and semantically sound sentences than with any other type of sentence. These findings are in line with previous research using the same type of task, since these sentences neither represent a challenge from the linguistic nor from the EC point of view (Bialystok, 1986b). Therefore, as it was discussed in the previous chapter, our results do not reveal a negative effect of bilingualism when evaluating the linguistic and metalinguistic abilities of the PWS population. Furthermore, our results do not reflect the BA found by Bialystok and collaborators for TD bilingual children but are in line with the behavioral results obtained by Moreno et al. (2010) for TD adults when using a similar task. According to these authors, the fact that adults, unlike children, have already attained linguistic mastery makes it difficult to compare these two populations, since children are still in the process of developing their linguistic skills. This proposal seems to imply that adults would principally rely on their linguistic knowledge to solve this type of task while children, due to their having less linguistic expertise, would rely more on non-linguistic factors, as for example their EC abilities. It is in this context where a BA seems to be predicated (see Friesen & Bialystok, 2012 for a similar approach when discussing Hermanto et al. 2012 results for bilinguals with different proficiency levels in each of their languages). The majority of our participants were adolescents or young adults with an age equivalent range for receptive vocabulary between 7;5-17;20 for monolingual speakers and between 7;5-16;3 for bilingual speakers. This, together with the fact that the majority of them are following or have followed an inclusive education program, leads us to believe that our participants would not be at an emerging linguistic knowledge stage, as it would be the case for TD children from previous studies (Bialystok, 1986b; Bialystok et al., 2014; among others), but at a more advanced developmental level. More data and the inclusion of two control groups consisting of TD monolingual and bilingual speakers would be needed in order to confirm the exact linguistic development stage of our participants. Nonetheless, at this point, it seems reasonable to assume that, during their school years, both monolingual and bilingual speakers have worked on and have developed their linguistic experience comparably and this would make them more similar to TD adults than to the TD children, at least at a passive level of linguistic knowledge and with the linguistic measures used in this dissertation.

When evaluating the role of the linguistic dimension, our results did not reveal better linguistic or metalinguistic abilities when judging sentences focused on order rather than on morphological linguistic phenomena, or vice versa. However, despite not finding significant differences, descriptive data in Table 17 suggests that monolingual speakers, but not bilinguals, seemed to show higher accuracy rates when judging ungrammatical and meaningful sentences focused on order rather than on morphology. At this point it is premature to draw conclusions, especially when inferential statistics have not revealed significant differences, and when it is obvious that more data is needed. However, these results could be helpful for designing future research experiments, since it would be interesting to explore with different experimental tasks and methodologies whether morphology is actually especially challenging for these individuals, as it has been proposed for other non-TD populations (see Paradis & Govindarajan, 2018 for an overview).

When analyzing the effect of individual differences in terms of the three standardized measures included in this study, results only suggested the nonverbal IQ abilities as a significant predictor when judging sentences in experiment 2. However, contrary to our prediction, higher nonverbal IQ abilities resulted in overall lower accuracy rates. This finding is striking and future research with more data and with different participants' cognitive profiles should explore the role of this variable more exhaustively to determine whether this finding is anecdotical or recurring. Likewise, the effect of the nonverbal IQ predictor should be explored in interaction with the Experimental Condition variable, something that we were not able to do here due to the limited data available. This would allow us to have a better understanding on how the nonverbal IQ abilities shape the linguistic and metalinguistic abilities of monolingual and bilingual speakers with PWS. We could venture to speculate that it could be the case that speakers with higher nonverbal IQ abilities (better logic skills) would perform worse when judging grammatically correct but meaningless sentences (less logic sentences). Also, it could be the case that the nonverbal IQ abilities play a different role in infancy and in adulthood since, while our results suggest worst outcomes when the nonverbal IQ is higher, Bialystok and Barac (2012) found a positive effect of the nonverbal IQ abilities when evaluating the metalinguistic abilities of English-Hebrew children, albeit with a metalinguistic task entailing low demands on EC and with a TD population.

In the case of bilingual speakers, no differences between languages were detected, which is surprising and deserves particular attention taking into account that they were predominantly Catalan dominant and that previous research using the same type of task, although with children, showed that, when bilinguals were tested in both their languages, they showed better results in their dominant language (Hermanto et al., 2012).

Experiment 3 consisted of different word-pairs manipulating the word length and the size of the objects and animals presented. In order to successfully complete the task, participants had to inhibit the semantic meaning of the words presented and, focusing on form, select the longest word of the pair (Bialystok, 1986a). Unlike in experiment 2, the results obtained were not very informative, since neither of the predictors included seemed to play a role in the results attained. Both monolinguals and bilinguals showed a very high accuracy rate (>85%) in all conditions independently of animacy. These results suggest, somehow, a ceiling effect, which implies that the task was very easy for this population given their literacy and their linguistic experience. However, it must be noted that when administering the task, longer times and more doubtful answers were perceived for incongruent items than for congruent or neutral ones. Thus, the next step is to adapt this experiment to an online task that would allow us not only to obtain accuracy data but also processing data (RT). Both types of data are complementary and would offer us a more comprehensive view of whether incongruent items are actually more demanding in terms of processing cost, which would translate in longer RTs, and of whether bilinguals would show shorter RT, which would indicate better conflict solving abilities.

7. The narrative abilities of monolingual and bilingual speakers with Prader-Willi Syndrome

While in the two previous chapters we have analyzed the EC and metalinguistic abilities of monolingual and bilingual speakers with PWS, in this chapter we investigate the narrative skills of this population. As we have indicated in chapter 3, research focusing on the linguistic abilities of the PWS population is very limited, essentially restricted to monolingual speakers and, with the exception of certain language-based standardized measures, it has mainly focused on these individuals' narrative macrostructural abilities.[64] In this chapter, we go one step further and investigate the narrative production of the PWS population at both the macrostructural and the microstructural (research question 3) level to shed light on whether monolinguals and bilinguals with PWS compare and whether dual speakers show similar effects of bilingualism to those observed for TD and other non-TD individuals (see sections 2.1 and 2.2 for a review). Also, as was the case in the two previous chapters, we are also interested in obtaining information about the role of the nonverbal IQ, receptive vocabulary, and sentence recall abilities in the narrative production process at both the macrostructural and the microstructural levels (research question 4), as well as in whether bilingual speakers show different narrative outcomes in both their languages (research question 5). Our hypotheses are the following: (i) bilinguals will outperform monolinguals at the macrostructural level but not at the microstructural level; (ii) if the three standardized measures included in this study play a role in the results, this is expected to be in terms of nonverbal IQ for the macrostructural abilities and in terms of receptive vocabulary for the microstructural abilities. Sentence recall abilities as a mixed measure of expressive language and

[64] The limited previous information about the morphosyntactic abilities of the PWS population came from conversational data (Lewis et al., 2002).

working memory could play a role in both narrative dimensions; and (iii) at the microstructural level bilinguals will show better outcomes in Catalan than in Spanish due to language dominance, but similar results at the macrostructural level.

This chapter is organized as follows. We first present the participants' sample for the narratives (what we have labelled experiment 4). In the second place we present the general procedure, transcription and coding scheme detailing how data was elicited and how utterances were computed as well as how words and different words were counted. We then present the different macrostructural and microstructural measures analyzed as well as the data analysis. Next, both dimension results are presented followed by the discussion of the results.

7.1. Participants' sample

A total of 13 participants (seven monolinguals and six bilinguals) were included in experiment 4. The exclusion of one monolingual and one bilingual from the original sample was due to the fact that their insufficient or unfocused narrative production was not deemed appropriate for evaluation. One of the bilingual speakers included in the between-group analysis was excluded from the within-group analysis because his/her Catalan narrative did not include enough material to be evaluated. Consequently, the comparative analysis of the narrative abilities among bilingual speakers included five participants. In Table 26 we present the nonverbal IQ, receptive vocabulary, and sentence recall abilities of the 13 participants of this experiment.

Table 26.Nonverbal IQ, receptive vocabulary and sentence recall abilities by group in Spanish (experiment 4). Standard scores (M=100, SD 15).

	Nonverbal IQ		Receptive Vocabulary		Sentence Recall	
	Monolinguals	**Bilinguals**	**Monolinguals**	**Bilinguals**	**Monolinguals**	**Bilinguals**
Mean	78.71	86.50	65.43	68.67	67.86	75.00
SD	6.55	5.82	17.87	15.71	15.77	20.98
Range	69.00-87.00	79.00-94.00	55.00-94.00	55.00-91.00	55.00-100.00	55.00-105.00

A mixed-design ANOVA with Standardized Test (Nonverbal IQ, Receptive Vocabulary, and Sentence Recall) as the within-subjects variable and Group (Monolingual vs. Bilingual) as the between-subjects variable was run to make sure that the exclusion of two participants did not compromise the groups' comparability. Results yielded a main effect of Standardized Test ($F(2, 22) = 3.82$, $p = .038$, $\eta_p^2 = 0.26$) but not of Group ($F(1, 11) = 1.58$ $p = .235$, $\eta_p^2 = 0.13$) and the interaction between both factors was not significant either ($F(2, 22) = 0.09$, $p = .914$, $\eta_p^2 = 0.01$). Post hoc pairwise comparisons with the Bonferroni correction revealed a significant mean difference between nonverbal IQ and receptive vocabulary abilities ($p = .029$, $d = 0.85$), as with the complete sample. In light of these results, we would like to conclude that the comparability between groups is maintained with these participants' sample.

Table 27.Receptive vocabulary and sentence recall abilities by language (experiment 4). Standard scores (M=100, SD 15).

	Receptive vocabulary		Sentence recall	
	Spanish	**Catalan**	**Spanish**	**Catalan**
Mean	71.40	96.40	79.00	78.00
SD	15.88	19.22	20.74	19.88
Range	55.00-91.00	67.00-118.00	55.00-105.00	55.00-105.00

Similarly, the exclusion of two bilingual speakers for the within-group analysis did not change the overall bilinguals' linguistic characteristics in both Spanish and Catalan described in section 4.3 (see Table 27 for descriptive data).

A repeated-measures ANOVA with Test (Receptive Vocabulary and Sentence Recall) and Language (Spanish and Catalan) as the within-subjects variables for bilingual speakers revealed a significant main effect of Language ($F(1, 4) = 16.36$, $p = .016$, $\eta_p^2 = 0.80$) and a significant interaction between Test and Language ($F(1, 4) = 30.45$, $p = .005$, $\eta_p^2 = 0.88$). A main effect of Test was not detected ($F(1, 4) = 0.17$, $p = .705$, $\eta_p^2 = 0.04$). Post hoc pairwise comparisons with the Bonferroni correction for the Test × Language interaction showed better receptive vocabulary abilities in Catalan than in Spanish ($p = .001$, $d = 2.95$). Thus, these results replicate the ones including the complete sample and show that bilinguals had a higher marginal mean in Catalan, and that they performed better in their dominant language at the receptive vocabulary level.

7.2. Procedure, transcription and coding

The 13 participants of the between-group analysis were asked to narrate Mayer's (1967) picture book *A boy, a dog and a frog* in Spanish. Participants were told the title of the story and were given a hard copy of the book. After this, they were asked to look at all the images carefully in order to be prepared to later narrate the story depicted. There was no time restriction and when participants felt ready, they started narrating the story. Bilingual speakers, in a different experimental session (see section 4.5 for the detailed task administration procedure), were asked to narrate Mayer's (1969) picture book *Frog, where are you?* in Catalan following the same procedure described for Spanish. Narratives were audio recorded. In order to avoid a practice effect, bilinguals were asked to narrate two different but intrinsically related stories, since *Frog,* *where are you?* is the continuation of *A boy, a dog and a frog* and both picture books included a similar number of images and events and the three main characters. The Frog Story picture books have been extensively used within the field to elicit narratives from individuals with different characteristics (see Slobin, 2005; Strömqvist & Verhoeven, 2004 for overviews).

The narrative corpora in both Spanish and Catalan were transcribed by the author of this dissertation, which is a Spanish-Catalan bilingual. Oral narratives were transcribed orthographically following CHILDES' (Child Language Data Exchange System) CHAT (Codes for the Human Analysis of Transcripts) conventions (MacWhinney, 2000). The macrostructural analysis of the narrations produced were evaluated by two independent Spanish-Catalan bilingual raters who had a background in linguistics and who were asked to follow the *Narrative Scoring Scheme (NSS)* (Heilmann et al., 2010). A detailed description of this protocol is provided when presenting the data analysis at the macrostructural level. As for the microstructure, the analysis is based on the participants' Mean Length of Utterance (MLU), which is considered a language development measure reflecting overall morphosyntactic complexity (R. Brown, 1973), and on the Type-Token Ratio (TTR)[65], which has been suggested to be a lexical variability measure (W. Johnson, 1944). Despite the fact that Spanish and Catalan are highly inflected languages, the MLU was computed by words and not by morphemes. The main aim of this microstructural measure was to have a glimpse of these individuals' morphosyntactic development and, taking into account the fact that previous research has found a high correlation between both MLU computations (e.g., Parker & Brorson, 2005), we decided to operationalize the MLU computation by words. Future research centered on a more in depth study of these individuals' morphosyntactic abilities should consider a more fine-grained analysis depending on the research objectives. In order to compute

[65] The TTR "expresses the ratio of different words (types) to total words (tokens) in a given language sample" (W. Johnson, 1944, p. 1).

the MLU accurately, together with the TTR, we have paid special attention to define what it is that we consider an utterance, how we have counted words and the decisions we have taken to define what counted as different words. The decisions taken to address these three issues are detailed next.

7.2.1. Utterance delimitations

As Turell and Moyer (2008, p. 201) points out, there is not a clear and consensual definition in the literature of what an utterance is and it is ultimately the researchers' decision to determine its boundaries taking into account the idiosyncrasies and main objectives of their research. Taking this into account, the criterion we have followed to determine utterances is the one proposed by MacWhinney (2000, p. 59), which defends that "utterances can include main clauses with associated depending clauses, but they should not include multiple main clauses". Therefore, independent clauses were transcribed in different lines (i.e., different utterances), as in (15), and dependent clauses were transcribed in the same line (i.e., one utterance) as in (16). Clauses with elided verbs such in (17) were considered independent clauses. Following Schneider, Dubé and Hayward's (2005) proposal, conjunctions such as *porque/perquè* 'because', *aunque/encara que* 'although'[66] and similar ones were considered to introduce dependent clauses, while conjunctions as *y/i*[67] 'and', *pero/però* 'but' and *entonces/llavors* 'then' were considered to introduce independent clauses. Also, following these authors, direct speech including a clause was transcribed as an independent utterance, as in (18), while indirect speech or direct speech not including a clause were transcribed as dependent clauses, as in (19).

[66] A clause including *aunque/encara que* '*although'* with concessive value is considered a dependent clause while when used with adversative meaning is considered an independent clause as synonym to *pero/però* 'but'.
[67] When two words are presented together separated by slash [/] the first word is in Spanish and the second one is the Catalan translation equivalent.

(15) PAR: el niño y el perro se marchan.
'The boy and the dog left'.
PAR: y la rana se queda triste.
'and the frog is sad'.

(16) PAR: el niño y el perro no sabían que la rana tenía bebés.
'the boy and the dog did not know that the frog had babies'.

(17) PAR: el niño cayó al agua.
'the boy fell into the water'.
PAR: y el perro también (cayó).
'and the dog (fell) too'.

(18) PAR: la rana dice +"/.
'the frog says'
PAR: +" me voy a vengar.
'I will take revenge'.

(19) PAR: el chico dice adiós.
'The boy says goodbye'.

Following what seems to be standard procedure (see Schneider et al., 2005), all production unconnected to the narration, as well as very short utterances signaling the end of the task, when included (e.g.,. *Ya está/ja està* 'That's it') were not included in the transcription to avoid an effect on the computation of the MLU and TTR.

7.2.2. Count word conventions

As previously stated, transcription was not phonological but orthographic and, as such, word count was strictly based on orthographic criteria. Phonetic and/or phonological idiosyncrasies were not included (a phonetic/phonological analysis is beyond the scope of this dissertation) and morphosyntactic errors were maintained to keep the participants production genuine. The decisions taken as to how to count words are detailed next. Contractions such as *del* (*de* + *el* ‘of +the$_{\text{MASC}}$’) or *al* (*a* + *el* ‘to +the$_{\text{MASC}}$’) in Spanish and Catalan were considered one word (unless the participant “broke” the contraction and specifically produced *de el* or *a el*) as well as enclitic pronouns in non-finite verbs like in *cogerlo* (*coger* + *lo* ‘*to catch it*$_{\textit{MASC-SING}}$’) in Spanish or *mirar-lo* (*mirar* + *lo* ‘to look at *it*$_{\textit{MASC-SING}}$’) in Catalan. However, when the clitic preceded the finite verb as in *lo cojo* ‘I$_{\text{NULL PRONOUN}}$ it$_{\text{CLI.MASC.SING}}$ take$_{\text{VERB}}$’ (‘I take it’) in Spanish or in *el miro* ‘I$_{\text{NULL PRONOUN}}$ it$_{\text{CLI.MASC.SING}}$ look at$_{\text{VERB}}$ (I look at it) in Catalan, it was counted as two independent words because they are two orthographic words. We are aware of the debate that treating proclitic and enclitic pronouns differently can raise but, for the sake of consistency, the orthographic criterion was strictly applied. Compound nouns such as, for instance, *cazamariposas* ‘butterfly net’ in Spanish were considered a single word. It is important to highlight that for comparability purposes, in those cases where Spanish and Catalan differ in how to compute orthographic words, the Spanish criterion prevailed (Spanish is the target language for between-group comparisons) and it was applied to Catalan to ensure, to the maximum extent possible, comparable MLU and TTR measures. The reason to do this was not to underestimate the morphosyntactic development and lexical variability of bilingual speakers when narrating in Catalan, since it needs to be noted that given the Catalan orthography, the exact same clauses between Spanish and Catalan could have been computed shorter in Catalan than in Spanish and this fact could have had important implications for the computation of both microstructural measures. For example, a Spanish clause like *el árbol es grande* (the$_{\text{DET.MASC.SING.}}$ tree$_{\text{NOUN.MASC.SING.}}$ is$_{\text{VERB.COP}}$ big$_{\text{ADJ.INV.}}$) has four orthographic words, however, the Catalan equivalent *l’arbre és gran* has only three. This is due to the fact that strictly orthographically speaking the Catalan determiner phrase is considered only one word by including the compulsory determiner elided form with its corresponding apostrophe.[68]

[68] The masculine singular determiner *el* in Catalan changes to *l’* when it precedes a noun starting with a vowel (excluding *i* or *u* functioning as consonants) or mute *h*.

If in this specific case, and in other similar ones, we would have not treated Spanish and Catalan the same way, we would have computed exactly the same clause (determiner + noun + copular verb + adjective) differently, and this is something we wanted to avoid to ensure, to the maximum extent possible, comparability among languages.

7.2.3. Different word conventions

Similarly to the importance of delimiting utterances and word boundaries for the computation of MLU and the total number of tokens, it was also necessary to delimit what were considered different types of words in order to compute the TTR. The general criterion was that every word used was a type except repeated words and those words with different forms but the same "meaning" (allomorphs).[69] In Catalan allomorphs are more common than in Spanish and when included they were coded as the same type. For example in a Catalan clause as *El nen va veure l'arbre* 'The boy saw the tree' both *el* and *l'* were computed as the same type (*el*-determiner). On the other hand, in case of polysemic words, independently of these being lexical or functional words, they were coded and computed as different types. For example in a clause like *la rana sigue al niño y él la invita a bañarse con él* 'the frog follows the kid and he invites her to take a bath with him' both *la* were coded as different types: (1) as a la_d (feminine and singular determiner) and (2) as a la_c (feminine and singular accusative clitic pronoun). Same words with different number and/or gender inflection were considered different words. For example, *rana* 'frog' and *ranas* 'frogs' as well as *la* 'The$_{\text{DET.FEM.SING.}}$' and *el* 'The$_{\text{DET.MASC.SING.}}$' counted as different types. Similarly, different inflection in person and/or number of the same verb counted as different types. For instance, *saltamos*, first person plural used as present 'we jump-present'; *saltamos*, first person

[69] We are referring to free morphemes. A morphological analysis including bound morphemes is beyond the scope of this dissertation.

plural used as preterite ‘we jumped-past’; and *saltan*, third person plural used as present ‘they jump’ (and the rest of the verbal forms with different person or tense/aspect/mood markers or meanings) counted as different words. Following this rationale, in those cases in which the same verbal form was used to refer to different persons, as for example the 1st and 3rd singular person of the imperfect tense, the different verbal forms were counted as different types. In code-switching cases, if a bilingual speaker referred to the same reality in both languages, as for example when using *rana* (Spanish) and *granota* (Catalan) to refer to the frog, both words were counted as two different types.

7.3. *Measures and data analysis: macrostructure and microstructure*

Macrostructural measures and data analysis

The 18 narratives (13 in Spanish and five in Catalan) were evaluated by two trained independent raters following the NSS conventions. Both raters were Spanish-Catalan bilinguals with a solid linguistic and pedagogical background. Narratives were presented randomized by language and raters were blind to the participants’ characteristics. Catalan narratives produced on the basis of the *Frog, where are you?* story were evaluated following the NSS scoring rubric for this picture book (SALT Software LCC, 2016). In the absence of an specific NSS scoring rubric for *A boy, a dog and a frog*, Spanish narratives for this picture book were evaluated with Bowers’ (2012) scoring rubric proposal. The NSS assesses the participants’ macrostructural narrative abilities evaluating the seven characteristics detailed in Table 28.

Table 28. Summary of the NSS evaluation criteria (SALT Software LCC, 2017).

Introduction	Adequate character and setting portrayal.
Character Development	Reference to both main and supporting characters through the development of the plot.
Mental and Emotional States	Use of language to describe the characters' thoughts and emotions.
Referencing/Listener Awareness	Use of clear and unambiguous referents and antecedents.
Conflict/Resolution and Event/Reaction	Accurate description of the different conflict/resolutions and event/reactions depicted in the story.
Cohesion	Information presented coherently and cohesively.
Conclusion	Adequate story ending.

Following the NSS protocol, each characteristic is evaluated according to a 0-5[70] scale where, in the authors' terms, five means "Proficient characteristics", three means "Emerging" and one means "Minimal/Immature". Two and four are two undefined values which are used as the judges' criteria to mark in between states among the three consolidated states previously presented. The sum of the different characteristic scores is used as a holistic narrative measure (see SALT Software LCC, 2017; for a detailed description of this narrative evaluation protocol).

It has been suggested that due to the NSS' somehow relative subjective evaluative component, interrater agreement may be lower than that found in more unbiased scoring systems (Heilmann et al., 2010, p. 161). Thus, in an attempt to overcome this potential subjective factor, the procedure followed in McFadden and Gillam (1996), which in turn is based on Myers' (1981) holistic scoring system, was followed to evaluate the interraters' agreements/disagreements. Accordingly, scores that differed by one point per characteristic and rater were not considered as disagreements. On the other hand, a two point difference (or higher) was considered a disagreement. The disagreements were annotated by the author of this dissertation and were

[70] Given that in the analysis we only included narratives with enough material to be evaluated, the score range was actually 1-5, since as described in the manual a score of 0 was reserved for severe deficiencies.

communicated to the two raters, who met and resolved them by consensus. Disagreements represented 14.29% of the cases (13 out of 91) in Spanish and a 5.71% in Catalan (2 out of 35). The interrater reliability for each characteristic and the whole narrative was then calculated with the Intraclass Correlation Coefficient (ICC), a widely-known measure to check for interrater reliability taking into account both the "degree of correlation and agreement between measures" (Koo & Li, 2016, p. 156). This index has been broadly used within the field to analyze interrater reliability when evaluating narrative abilities among TD and non-TD individuals (Estigarribia et al., 2011; Gonçalves et al., 2002, 2011; Zevenbergen et al., 2003; among others).

ICCs were computed using SPSS statistical package version 25 (IBM Corp., 2018) based on a mean-rating ($k = 2$), absolute-agreement, 2-way mixed-effects model. According to Koo and Li (2016, p. 158), a possible ICC interpretation is the following: below .5 = poor reliability, between .5-.75 = moderate reliability, between .75-.9 = good reliability and greater than .90 = excellent reliability. The standard cut point for interrater reliability analysis is normally stablished at .70, although this cut-off threshold has been open to debate (see Lance et al., 2006 for more details). The ICC values for each characteristic and whole narratives in both languages are presented in Table 29.

Table 29. Intraclass Correlation Coefficients for Spanish and Catalan narratives. Holistic and analytical scores.

	Spanish	Catalan
Holistic narrative evaluation	.94	.93
Introduction	.91	.91
Character Development	.78	.56
Mental and Emotional States	.97	.98
Referencing/Listener Awareness	.54	.89
Conflict/Resolution or Event/Reaction	.72	.83
Cohesion	.71	.71
Conclusion	.97	.89

In the Spanish narratives all ICCs values were above the recommended .70 cut-off point except for the *referencing/listener awareness* characteristic. Thus, given the low reliability of this characteristic, and adopting a conservative approach, as in Finestack et al. (2012), this characteristic was not further analyzed and the holistic narrative evaluation score was recomputed (maximum score = 30) omitting this characteristic's scores (interrater ICC = .95). In Catalan narratives, only the Character development characteristic did not reach the minimum recommended threshold of .70. As a consequence, when comparing the macrostructural narrative abilities of bilingual speakers between languages, the comparison analysis did neither include *referencing/listener awareness* nor *character development* characteristics. The holistic narrative evaluation score in both languages was therefore recalculated (maximum score = 25) without these two characteristics' scores (interrater ICC $\geq$ .93 in both languages). The final scores entered into the analysis were the average between both raters per participant and characteristic (analytical evaluation) or whole narrative (holistic evaluation).

We performed a between-group analysis and a within-group analysis for bilingual speakers. In the between-group analysis we analyzed both groups' macrostructural abilities in Spanish at a holistic and an analytical level and in the within-group analysis we scrutinized the macrostructural narrative abilities of bilingual speakers in both Spanish and Catalan. We first present the between-group analysis followed by the within-group analysis.

Between-group analysis (monolingual vs. bilingual speakers)

Both holistic and analytical narrative scores were analyzed separately using a multiple regression analysis (*lm* model) and a linear mixed model analysis (*lmer* model), respectively. For the holistic narrative analysis, we run a multiple regression analysis with the Global Score as outcome variable and with Group (-0.5 for Monolinguals vs. 0.5 for Bilinguals), Nonverbal IQ, Receptive Vocabulary and Sentence Recall (all three *z*-scored) as independent variables. As a multiple regression analysis was “simpler” than a linear mixed model analysis in the sense that participants only contributed one data point, we run the analysis including all the predictors of interest at once instead of breaking the analysis in two phases, as we did in the previous analyses. This model was then submitted to a simplification process by removing one predictor at a time. The original model was then compared to its simplified version and if its AICc value improved, this model was kept and further simplified. Ultimately, the model with the lowest AICc was selected as the best-fitting model.

When analyzing the narrative abilities from an analytical perspective, we first run a linear mixed model analysis with Analytical Score as outcome variable and with Narrative Characteristic (Introduction, Characters, Mental, Conflict, Cohesion, Conclusion) in interaction with Group (-0.5 for Monolinguals vs. 0.5 for Bilinguals) as fixed effects. The random effects structure included a by-participant intercept and a slope for the Narrative Characteristic by participant.[71] The simplification process was the same as the one applied in chapters 5 and 6, which is described in detail in chapter 4.

Within-group analysis (Spanish vs. Catalan among bilingual speakers)

The holistic narrative abilities in both languages were compared using a paired sample *t*-test and the analytical narrative abilities were compared using a repeated-measures ANOVA.

[71] As it will be further discussed in the results section, given that the Narrative Characteristic predictor did not improve the model-fit, we did move forward to phase 2 (i.e., inclusion of the three standardized measures as fixed effects).

Microstructural measures and data analysis

The MLU and the TTR indexes were computed on the Spanish and Catalan corpora running two Computerized Language Analysis (CLAN) commands: MLU and FREQ, respectively. With these measures, we then performed two different statistical analyses as in the previous section: a between-group analysis and a within-group analysis.

Between-group analysis (monolingual vs. bilingual speakers)

The MLU and TTR data were analyzed separately using a multiple regression analysis (*lm* model) with the MLU and TTR scores as outcome variables. In both analyses, Group (-0.5 for Monolinguals vs. 0.5 for Bilinguals), Nonverbal IQ, Receptive Vocabulary, and Sentence Recall (all three *z*-scored) were included as independent variables. Additionally, the Narrative Length (number of words)[72] variable was entered into the analysis in interaction with Group to check for a potential narrative length effect on the participants' score.[73] This variable is especially relevant for the TTR analysis, since it has been argued that the oral/written production length can somehow influence the TTR outcomes, i.e., longer productions may increase the chances of obtaining lower TTR scores (Fergadiotis et al., 2015; Mellor, 2011; among others). The modelling process replicates the one presented in the previous section.

Within-group analysis (Spanish vs. Catalan among bilingual speakers)

Bilingual speakers' MLU and TTR measures in both Spanish and Catalan were compared using paired sample *t*-test.

[72] The number of words is equivalent to the number of tokens.
[73] Although being monolingual or bilingual does not seem to a be significant predictor of the narrative length, the Narrative Length variable was entered into the model in interaction with Group to make sure a potential interaction between these two variables was not overlooked when modelling the MLU and the TTR data.

7.4. Results

Descriptive data is presented first followed by the inferential statistical analyses for both the between-group and the within-group analyses at the macrostructural level first and the microstructural level next.

Results at the macrostructural level

Between-group analysis (monolingual vs. bilingual speakers)

Descriptive data in Table 30 shows that both groups demonstrate certain shortcomings when it comes to narrating, since practically all the mean scores are below 3 (i.e., emerging narrative state). Overall, participants show similar narrative abilities in all the narrative characteristics under analysis and this is the case independently of their being monolinguals or bilinguals. However, it is important to highlight that, even though the outcomes are similar, bilingual scores are slightly higher across the board, especially for the *character development* and *conclusion* characteristics.

Table 30. Mean holistic and analytical narrative scores by group in Spanish. SD in parentheses.

	Monolinguals	Bilinguals
Holistic narrative evaluation[a]	15.07 (5.47)	17.92 (3.28)
Introduction[b]	2.57 (0.89)	2.83 (0.98)
Character Development	2.86 (0.90)	3.42 (0.38)
Mental and Emotional States	2.29 (1.29)	2.83 (0.98)
Conflict/Resolution or Event/Reaction	2.43 (1.10)	2.92 (0.49)
Cohesion	2.43 (0.61)	2.75 (0.52)
Conclusion	2.50 (1.47)	3.17 (1.17)

[a] The maximum score for the holistic narrative evaluation is 30.
[b] The maximum score per characteristic is 5.

In order to confirm these insights, we first analyzed the participants' Holistic Narrative score with a multiple regression analysis including Group, Nonverbal IQ, Receptive Vocabulary, and Sentence Recall as predictors. The best-fitting model (see Table 31) only included the Receptive Vocabulary variable, which explained a 13% of the data variance but it did not contribute significantly to the model ($F(1, 11) = 1.72$, $p = .217$). These results indicate that none of the predictors included in this study seems to play a significant role in predicting the participants' holistic narrative score.

Table 31. Model parameters for the best-fitting model. Dependent variable: holistic narrative score.

	Estimate	SE	*t*	Pr(>|*t*|)
Intercept	16.39	1.25	13.07	< .001*
Receptive Vocabulary	1.71	1.30	1.31	.217

Receptive Vocabulary was *z*-scored

We then modelled the Analytical Narrative score outcome variable starting with the maximal model presented in section 7.3. The model converged with the by-participant intercept and with the Narrative Characteristic in interaction with Group as fixed effects. The best-fitting model was the one including the participant intercept and the fixed effect of Group (estimate = 0.47, $SE = 0.43$, $t = 1.11$, $p = .291$)[74], which did not reveal differences between monolingual and bilingual speakers. These results provide evidence that the repeated measure variable did not improve the model fit, which is interpreted as not playing a significant role in the results obtained. Thus, we did not continue further with this analysis, since proceeding with an analysis not including the different narrative characteristic would overlap with the holistic narrative analysis previously detailed.

[74] Fixed effects: intercept (estimate = 2.75, $SE = 0.21$, $t = 12.87$, $p < .001$).
Random effects: participant intercept ($SD = 0.72$).

Within-group analysis (Spanish vs. Catalan among bilingual speakers)

Descriptive data in Table 32 suggest that bilingual speakers show comparable macrostructural narrative abilities in Spanish and Catalan at both the holistic and the analytical level.

Table 32. Mean holistic and analytical narrative abilities of bilingual speakers in Spanish and Catalan. SD in parentheses.

	Spanish	Catalan
Holistic narrative evaluation[a]	15.10 (2.88)	14.00 (4.43)
Introduction[b]	3.20 (0.45)	3.20 (1.04)
Mental and Emotional States	2.80 (1.10)	2.50 (1.41)
Conflict/Resolution or Event/Reaction	3.00 (0.50)	2.80 (0.76)
Cohesion	2.90 (0.42)	2.80 (0.57)
Conclusion	3.20 (1.30)	2.70 (1.15)

[a] The maximum score for the holistic narrative evaluation is 25.
[b] The maximum score per characteristic is 5.

In order to verify this, we first run a repeated measures ANOVA with Narrative Characteristic (Introduction, Mental, Conflict, Cohesion, Conclusion) and Language (Spanish and Catalan) as the within-subjects variables. The Mauchly's test indicated that the sphericity assumption was not met for the Narrative Characteristic variable ($\chi 2(9) = 19.80$, $p = .047$) and consequently the degrees of freedom were adjusted using the Greenhouse-Geisser correction ($\varepsilon = .51$). Results did not yield either a main effect of Narrative Characteristic ($F(2.03, 8.13) = 0.77$, $p = .495$, $\eta_p^2 = 0.16$), or of Language ($F(1, 4) = 0.24$, $p = .650$, $\eta_p^2 = 0.06$). The interaction between both factors was not significant either ($F(4, 16) = 0.31$, $p = .869$, $\eta_p^2 = 0.07$). These results reflect that bilingual speakers show comparable outcomes for all the narrative characteristics in both Spanish and Catalan. Similarly, a paired sample *t*-test revealed similar holistic narrative abilities in both languages ($t(4) = 0.49$, $p = .650$, $d = 0.22$).

In summary, results show that both monolingual and bilingual speakers have similar macrostructural narrative abilities and that they do not display shortcomings or strengths for different narrative characteristics included in the NSS scoring system, but similar outcomes across the board. Bilingual speakers display comparable results at the microstructural and the macrostructural level in both Spanish and Catalan.

Results at the microstructural level

Between-group analysis (monolingual vs. bilingual speakers)

Descriptive data in Table 33 suggest that monolingual and bilingual speakers show similar MLU and TTR scores. However, despite overall similarities both measures show different tendencies: higher MLU for monolinguals and higher TTR for bilinguals. In terms of narrative length, it seems that monolingual speakers would be more productive than bilingual speakers, since on overage they produce narratives 75 words longer. However, a simple regression analysis indicated that group was not a significant predictor of the narrative length ($F(1, 11) = 1.43$, $p = .257$).

Table 33.Number of words (= tokens), types, MLU and TTR by group in Spanish.

	Number of Words (=Tokens)		Number of different Types		MLU		TTR	
	Monolingual	**Bilingual**	**Monolingual**	**Bilingual**	**Monolingual**	**Bilingual**	**Monolingual**	**Bilingual**
Mean	338.29	262.83	101.29	89.83	8.20	7.20	0.31	0.34
SD	138.71	72.20	41.48	22.87	1.31	1.49	0.05	0.02
Range	144-541	156-338	54-170	56-109	6.35-10.06	4.73-8.73	0.22-0.38	0.32-0.36

In order to determine whether monolingual and bilingual speakers showed comparable results and whether the narrative length, the nonverbal IQ, the receptive vocabulary, and the sentence recall abilities might play a role in predicting the MLU and TTR scores, we ran two multiple regression analyses as detailed in the analysis section. We present first the MLU results followed by the TTR ones.

The best-fitting model for the MLU variable outcome (see Table 34**Table**), which explained 17% of the data variance, only included the fixed effect of Narrative Length and did not play a significant role predicting the MLU score (*F*(1, 11) = 2.24, *p* = .163). Thus, according to these results, we can conclude that both groups presented a similar morphosyntactic complexity development and that none of the other four predictors included in the analysis played a role in the participants' MLU scores.

Table 34. Model parameters for the best-fitting model. Dependent variable: MLU.

	Estimate	**SE**	***t***	**Pr(>|*t*|)**
Intercept	6.19	1.10	5.62	< .001*
Narrative Length	0.01	0.00	1.50	.163

The best-fitting model for the TTR variable outcome (see Table 35) included Group, Nonverbal IQ, and Sentence Recall as predictors. The model explained 69% of the data variance (adjusted R2 = 59%), which is pretty high for linguistic data (see Winter, 2019, p. 106), and the inclusion of the aforementioned predictors played an important role in predicting the TTR scores (*F*(3, 9) = 6.71, *p* = .011). More precisely, results indicated that, when keeping all the predictors constant, a significant difference between groups arose revealing a higher lexical variability for bilingual speakers (*t* = 0.96, *p* = .003). Also, results evidenced that, when holding all the predictors constant, an increase in the Nonverbal IQ (*t* = -2.87, *p* = .018) and sentence recall abilities (*t* = -2.93, *p* = .017) resulted in a decrease of the TTR score. These results are puzzling and we will come back to this issue in the discussion section. The standardized beta values (β) show that the fact of being monolingual or bilingual is the central predictor of the model followed by the nonverbal IQ abilities and by the sentence recall abilities in this order.

Table 35. Model parameters for the best-fitting model. Dependent variable: TTR.

	Estimate	SE	*t*	β	Pr(>|*t*|)
Intercept	0.33	0.01	43.28		<.001*
Group	0.08	0.02	4.09	0.96	.003*
Nonverbal IQ	-0.03	0.01	-2.87	-0.66	.018*
Sentence Recall	-0.02	0.01	-2.93	-0.57	.017*

The Group predictor is contrast-coded (monolinguals, coded as -0.5 vs. bilinguals, coded as 0.5) and the Nonverbal IQ and Sentence Recall predictors are *z*-scored.

Within-group analysis (Spanish vs. Catalan among bilingual speakers)

Descriptive data in Table 36 show that bilingual speakers display similar MLU and TTR outcomes in both languages, which was confirmed by inferential statistics ($t(4) = 1.85$, $p = .138$, $d = 0.83$; $t(4) = -0.61$, $p = .573$, $d = -0.28$; respectively). When it comes to the length of the narratives, results suggest that bilingual speakers may be more productive in Catalan than in Spanish, however this difference was not statistically significant ($t(4) = 1.70$, $p = .165$, $d = 0.76$).

Table 36. Number of words (= tokens), types and MLU and TTR indexes in Spanish and Catalan for bilingual speakers (experiment 4).

	Number of Words (=Tokens)		Number of different Types		MLU		TTR	
	Spanish	Catalan	Spanish	Catalan	Spanish	Catalan	Spanish	Catalan
Mean	284.20	355.80	96.60	124.40	7.70	7.05	0.34	0.35
SD	55.61	43.92	17.62	16.77	0.97	1.57	0.02	0.05
Range	192-338	304-424	68-109	100-147	6.37-8.73	5.68-9.34	0.32-0.36	0.28-0.42

In summary, the MLU and TTR outcomes seem to be sensitive to different predictors. While neither Group nor any of the three standardized measures included in this study appeared to predict the MLU scores, the fact of being monolingual or bilingual together with the participants' nonverbal IQ and the sentence recall abilities predicted, albeit in the opposite direction to what we expected, the participants' TTR scores. As was the case with the metalinguistics tasks, bilingual speakers exhibited comparable results in both Spanish and Catalan.

7.5. Discussion

The main aim of this chapter was to compare the narrative abilities of monolingual and bilingual speakers with PWS. To address this, we elicited oral narratives using the Frog Stories series and we analyzed the participants' holistic and analytic narrative macrostructural and microstructural abilities using the NSS scoring system and the MLU (in words) and the TTR indexes, respectively. Overall, the results did not evidence that being monolingual or bilingual was a significant predictor of the different measures under study, although an exception for the TTR results was found. In this case, bilingual speakers, while maintaining the nonverbal IQ and the receptive vocabulary abilities constant, produced significantly richer narratives in terms of lexical variability. These findings, which do not show a detrimental effect of bilingualism, are in line with the results of the two previous chapters and with the results of previous research centered on the study of bilingualism in individuals with DS (Feltmate & Kay-Raining Bird, 2008; Kay-Raining Bird et al., 2005; among others). Furthermore, the BA detected is also in line with previous research revealing some strengths of bilingual speakers for certain microstructural measures for DLD-SLI (Tsimpli et al., 2016) and ASD bilinguals (Peristeri et al., 2020), albeit with different measures. We will now elaborate on these findings.

Focusing first on the microstructural results, it may look surprising that the group showing richer narratives in terms of vocabulary is the bilingual one, since it has been argued that bilinguals normally show lower receptive vocabulary abilities (Bialystok et al., 2010; Bialystok & Luk, 2012; among others) than their monolingual peers. This type of research has traditionally put the bilinguals' receptive vocabulary abilities to the test using standardized tests or very controlled experimental tasks. However, in this case, participants were not evaluated on their passive vocabulary but on their expressive one and in a situation where they were in control of what they said and how they said it. Additionally, and with all due caution, taking into account the spontaneous nature of the task and the few constraints imposed, the bilinguals' better lexical variability outcomes could also be interpreted in terms of better bilingual linguistic creativity, since previous research has suggested better divergent thinking abilities for dual speakers (see Adesope et al., 2010). Alternatively, the BA found could be explained in terms of language mixing, since, as noted when detailing how different words were counted, the same lexical word in Spanish and Catalan in the same narrative counted as two different types. These cases, although occasional, are an idiosyncrasy of the bilingual narratives. Another factor to take into account is the proximity between Spanish and Catalan, two languages from the same family (Romance languages), which does not allow us to rule out the hypothesis of a positive crosslinguistic transfer effect at the lexical level benefiting bilingual speakers. In addition to all this, it is fair to acknowledge a possible effect of the TTR measure itself. We are aware that one of the main criticisms of the TTR index as a measure of lexical variability is its sensitivity to length, since longer productions can be penalized in terms of TTR scores (Fergadiotis et al., 2015; Mellor, 2011). Nonetheless, this pattern was not observed in our corpus, since the best-fitting model for the TTR results did not include the narrative length variable. Be that as it may, this possible positive effect of bilingualism should be interpreted cautiously and should be confirmed in future research with different indices assessing the lexical diversity/variability abilities, with different types of narrative elicitation methods and with different language pairings.

Focusing on the MLU outcomes, our results did not reveal group as a significant predictor. Now, the next step would be to investigate how TD and PWS individuals compare in terms of MLU. We do not have direct comparable data to carry out this contrast, but if we compare the mean MLU score of monolinguals (MLU = 8.20) and bilinguals (MLU = 7.20) with PWS in this chapter with the results obtained by García-Alcaraz (2018) from a TD adult English-Spanish bilingual when narrating exactly the same picture book (*A boy, a dog and a frog*) in Spanish, the PWS population does not seem to differ much from this TD individual (MLU = 7.84). Nonetheless, it needs to be stressed that, for this TD bilingual, Spanish was not his/her dominant language but the heritage language and when this same picture book was narrated in English (dominant language) the MLU increased (MLU = 11.10). If a similar pattern is found when comparing the MLU results of the TD and the PWS population in Spanish, these results would be in line with Lewis' (2002), who compared the MLU (in morphemes) of TD and PWS individuals using conversational data at different age ranges (from childhood to adulthood) and showed that the PWS individuals fell behind at all developmental stages.

In terms of the macrostructural abilities, our findings evidenced similar narrative shortcomings for both monolinguals and bilinguals across the board, since all the analytical scores were below 4 and the majority of them were below 3, which reflects an emerging narrative state. These results are in line with previous research evidencing that narration is a hard task for monolinguals (Garayzábal-Heinze et al., 2012; Lewis et al., 2002) and bilinguals with PWS (García-Alcaraz, 2018). However, once again, despite the fact that inferential statistics did not reveal a significant positive effect of bilingualism, descriptive data (see Table 30) showed higher mean scores for bilinguals for all the narrative characteristics as well as at a global level, which could be, somehow, pointing to a potential BA.

Spanish-Catalan bilinguals produced similar narratives in both their languages for the microstructural and the macrostructural measures. These results are in line with Fiestas and Peña (2004) and with Kunnari et al. (2016), who did not find differences between TD bilinguals' children languages when assessing their narrative macrostructural abilities. The absence of differences between languages at the macrostructural level is in line with Gagarina et al.'s (2016) hypothesis, which, as detailed in chapter 2, states that differences between bilinguals' language, if found, are expected at the microstructural but not at the macrostructural level, since the latter does not focus on specific language use but addresses a more cognitive dimension and these cognitive abilities would be easily transferrable between languages. At the microstructural level, it is noteworthy that, despite the fact that our bilingual speakers are predominantly Catalan dominant, they show similar outcomes for their non-dominant language, which aligns with Fiestas and Peña's (2004) results showing similar productivity outcomes for the two languages of TD Spanish-English bilingual children.

Lastly, regarding the role played by the nonverbal IQ, receptive vocabulary and sentence recall abilities in the results obtained, our findings only revealed as predictive factors the nonverbal IQ and the sentence recall abilities for the TTR outcomes. However, contrary to what we hypothesized, the effects of these two variables were negative, i.e., higher nonverbal IQ and higher sentence recall abilities resulted in a decrease of the TTR score. A priori these results seemed puzzling, but a demotivating effect could have been the cause for such results. When we presented the narrative task, some participants, especially those with higher functioning abilities, were not very excited about the idea of narrating a picture book, since they perceived it as a children's task, but they finally freely agreed to complete it. It is possible that some of them, albeit having been told that it was a common task used with both children and adults, may not have been very motivated and did not put a lot of effort on the task resulting in more repetitive and less elaborated narratives. Future studies assessing the narrative abilities of this population should find a more engaging elicitation method for older participants that somehow ensures the participants' stimulation and neutralizes a possible (de)motivation effect, which we suspect is the cause of the unexpected findings.

In summary, this chapter has revealed, overall, similar outcomes for both monolingual and bilingual speakers with PWS at the macrostructural and microstructural level and no differences between languages at these two levels for bilingual speakers. Moreover, in this case a significant BA has been detected when evaluating the participants' lexical variability abilities.

8. General discussion and conclusions

The main aim of this dissertation was to compare the EC, metalinguistic and narrative abilities of monolingual and bilingual individuals with PWS to shed light on whether the fact of speaking two languages does not result in negative effects for the PWS as has been proposed for other DD (Kay-Raining Bird et al., 2016; Paradis & Govindarajan, 2018). Additionally, we were interested in elucidating whether the potential beneficial effects of bilingualism suggested for the TD population, and more recently for the non-TD population in different contexts (see Chapter 2 for a review), showed in our data. To meet this goal, we administered four experimental tasks to eight Spanish monolingual speakers with PWS and to seven Spanish-Catalan speakers with the same syndrome. These tasks targeted abilities for which previous research has suggested a potential BA. Experiment 1 consisted of two EC tasks intended to compare monolingual and bilinguals' interference suppression abilities using the Flanker task (a non-language based task) and the Stroop task (a language driven task). Experiment 2 and experiment 3 evaluated the participants' metalinguistic abilities using a SJT and a WLJT, respectively. Experiment 2, beyond the assessment of the participants' metalinguistic abilities also evaluated their grammatical knowledge. Lastly, with the oral narratives, which were labeled as experiment 4, the participants' microstructural and macrostructural narrative abilities were evaluated. In addition to these experimental tasks, participants also completed different standardized tests intended to obtain a clearer picture of their nonverbal IQ abilities as well as their receptive vocabulary and sentence recall abilities. On the one hand, these three measures allowed us to have a better understanding of the participants' non-verbal and linguistic profile and, on the other hand, enabled us to determine the comparability of monolinguals and bilinguals with respect to those abilities. Beyond its main objective, this dissertation also investigated the global role of these three cognitive and linguistic measures on the different experimental tasks' results and whether bilingual speakers showed different linguistic abilities in Spanish and Catalan at the metalinguistic and narrative levels. To put the findings in perspective, this final chapter is organized as follows: in section 8.1 we recapitulate the main outcomes by answering the five research questions set out in chapter 4. In section 8.2 we succinctly discuss the effects of bilingualism on the PWS and we discuss the results obtained in the light of the BAH. Next, the contributions and implications of this dissertation as well as the limitations and future research recommendations are presented in section 8.3.

8.1. The cognitive and linguistic abilities of monolinguals and bilinguals with Prader-Willi Syndrome.

In chapter 4 we laid out five research questions which are discussed below.

<u>Research Question 1:</u> *How do bilinguals and monolinguals with PWS compare with respect to EC, in particular in relation to the interference suppression abilities?*

Based on previous research focused on bilingualism and genetic syndromes (see section 2.3 for a review), we did not anticipate negative effects of bilingualism for the PWS population. In fact, on the basis of the BAH and previous findings suggesting a BA at the EC level for bilinguals with ASD (Baldimtsi et al., 2016, 2020; Gonzalez-Barrero & Nadig, 2019b; Peristeri et al., 2020), if an effect of bilingualism was detected for the PWS population this was expected to be positive.

Our results clearly showed no negative effects of bilingualism. However, while descriptive results for RT data pointed to a general BA by showing overall faster RTs for bilingual speakers as well as smaller interference effects in both EC tasks and an smaller facilitation effect on the

Stroop task, the inferential statistical analysis either did not include the group variable into the best-fitting model or if included, it did not reveal significant differences between groups. Thus, neither instances of the BICA hypothesis (lower RTs for bilinguals in the incongruent condition) nor of the BEPA hypothesis (lower RTs for bilinguals across the board) were confirmed. Both monolingual and bilingual speakers showed higher RTs for incongruent trials when compared to neutral trials (interference effect) but similar RTs when comparing the congruent and neutral conditions (no facilitation effect). Similarly, accuracy data revealed a very high accuracy rate among participants independently of their being monolingual or bilingual speakers (> 90%). This high level of offline performance revealed not only that participants understood the task, which was also confirmed with the practice trials administered before the actual task, but also that they did not show impaired interference suppression abilities, or at least not with this type of measure and with these types of tasks.

Research question 2: *How do bilinguals and monolinguals with PWS compare with respect to metalinguistic abilities?*

Between the two metalinguistic tasks administered, the most revealing results were those of the SJT. The WLJT task, on the other hand, did not allow us to see much of the underlying metalinguistic abilities of PWS due to it being very easy for this population and leading to a ceiling effect. Therefore, we concentrate on the SJT task results to discuss the metalinguistic abilities of the PWS population.

As with the EC abilities, no negative effects of bilingualism were anticipated at the metalinguistic level and if differences between groups were detected these were anticipated to be positive for bilingual speakers, as in Gonzalez-Barredo and Nadig's (2017) study, where a BA was found when administering a metalinguistic task entailing high demands on EC to ASD individuals.

Taking into account previous research results assessing the metalinguistic abilities of the TD population within Bialystok's metalinguistic cognitive framework using a SJT (Bialystok, 1986, 1988; Bialystok et al., 2014; among others), it was hypothesized that if bilingualism played a similar role for the non-TD population to that suggested for the TD-population, bilingual speakers with PWS would be expected to outperform monolingual speakers when judging meaningless but grammatical sentences (essentially a metalinguistic measure) but not necessarily when judging ungrammatical but semantically sound sentences (essentially a linguistic measure). In fact, and due to language dominance, if a bilingual effect was detected for the latter sentences, we expected this effect to be negative.

Descriptive data revealed that, on average, bilingual speakers had a slightly higher score than monolingual speakers when judging sentences assessing essentially metalinguistic abilities, while the reverse pattern, with a rather higher difference, was observed when evaluating fundamentally grammatical knowledge. Nevertheless, as was the case when we evaluated the EC abilities, the inferential statistical analysis did not reveal differences between groups: neither a BA nor a bilingual disadvantage for any of the measures under evaluation, not even for the more linguistic ones. This finding is especially relevant since it is important to bear in mind that bilinguals were compared to their monolingual peers in their non-dominant language and, even if they were at a disadvantage, they did not show grammatical knowledge shortcomings when compared to Spanish monolingual speakers.

It is also important to emphasise that the best-fitting model for the SJT data did not include the linguistic dimension variable, which suggests that the distinction between order and morphology did not play a role in the results obtained. We interpret this as meaning that sentences manipulating both order and morphology entailed a similar degree of difficulty for this population.

<u>Research question 3:</u> *How do bilinguals and monolinguals with PWS compare with respect to narrative abilities at the macrostructural and microstructural levels?*

As in the previous two research questions, no detrimental effect of bilingualism was anticipated. Again, assuming the hypothesis of similar effects of bilingualism at the narrative level similar to those observed for the TD and non-TD population (see sections 2.1.3 and 2.2.3 for a review), differences between groups, if present, were expected at the narrative cognitive level, i.e., the narrative macrostructure (measured using the NSS scoring system), than at the language use level, i.e., the narrative microstructure (measured with the MLU and the TTR indexes). However, the results revealed the opposite pattern: comparative narrative shortcomings for monolingual and bilingual speakers at the macrostructural level but better lexical variability abilities for the bilingual group. In this case, bilinguals outperformed their monolingual peers possibly due to a positive crosslinguistic transfer effect facilitated by the similarity between Spanish and Catalan, both Romance languages. However, although with different measures, Tsimpli et al. (2016) and Peristeri et al. (2020) also reported benefits of bilingualism at the microstructural level for individuals with DLD-SLI and ASD, respectively. In the case of the MLU, no differences between groups were attested.

<u>Research question 4:</u> *Do participants' individual differences in terms of nonverbal IQ, receptive vocabulary and sentence recall abilities play a role in explaining some of the overall data variance observed when analyzing the participants' EC, metalinguistic and narrative ability results?*

The analysis of the participants' cognitive and linguistic characteristics revealed that overall both monolingual and bilingual speakers showed comparable nonverbal IQ, receptive vocabulary and sentence recall abilities. Nevertheless, both groups showed higher nonverbal IQ than receptive vocabulary abilities. Taking the receptive vocabulary scores as a proxy for verbal IQ (Bell et al., 2001; Hodapp & Gerken, 1999; Krasileva et al., 2017), our participants showed better nonverbal than verbal abilities, which, considering that most of them belonged to de DEL PWS subtype, constitute results that are in line with previous research (Copet et al., 2010; Dimitropoulos, Ferranti, et al., 2013). In the specific case of bilingual speakers, they showed better receptive vocabulary abilities in Catalan than in Spanish but similar outcomes for both languages with respect to the sentence recall scores. These results are not surprising if we take into account that bilingual speakers were mostly Catalan dominant; although it is important to exercise caution as the Catalan results, unlike the Spanish results, were not standardized.

When the global effect of these three participants' characteristics was analyzed to investigate whether better nonverbal IQ, receptive vocabulary, and sentence recall abilities led to overall better outcomes in the different experimental tasks administered, the results revealed that, in general, they did not play a significant role, with the exception of the SJT and the TTR results. In both cases the nonverbal IQ predictor was significant and in the case of the TTR measure so were the sentence recall abilities. Nevertheless, contrary to our hypothesis, which predicted that if these measures played a role, the effect would be positive, our findings showed the opposite: higher abilities lead to worse outcomes. As discussed in chapters 6 and 7, at this point our account of these findings are speculative and future research should determine whether these effects are truly a pattern. Our explanation for the negative effect of the nonverbal IQ in the SJT task is that these results could somehow be showing that individuals with higher nonverbal IQ scores, i.e., more logical; performed poorly when judging grammatically correct but semantically anomalous sentences, i.e., the less logical sentences. In other words, what we are proposing here is an interaction between the experimental condition variable and the nonverbal IQ scores and this is something that future research can easily (dis)prove with a wider sample that allows to explore this interaction with more robust data so that all the predictors can be introduced at once. In the case of the TTR results, the most feasible explanation that the data available allow us to consider is a demotivation effect: participants with higher nonverbal IQ and better sentence recall abilities perceived the narration task as a children's task not worth their effort. Future research should take this into consideration and look for more engaging alternative narrative elicitation tasks that could neutralize a potential demotivation effect.

These findings showed that, in general, the three standardized measures were not a significant predictor of the results obtained but when an effect was found the nonverbal IQ abilities emerged as the more relevant one. Other researchers have highlighted the relevance of IQ (either full, verbal or nonverbal) when assessing the EC (Chevalère et al., 2015) and the metalinguistic abilities (Jauregui et al., 2007) of the PWS population, although in these cases they have showed a more predictable relationship. Taking this into account, future research should pay special attention to the participants' (non)verbal IQ when analyzing and interpreting experimental and spontaneous data.

Research question 5: *Do bilingual speakers show comparable metalinguistic and narrative abilities in both Spanish and Catalan or do they perform better in Catalan, the predominantly dominant language?*

On the basis of language dominance, we anticipated that bilingual speakers would show similar outcomes in both of their languages for the less linguistic driven measures, i.e., in the grammatical but meaningless sentences in the SJT and in the macrostructural narrative analysis; and better outcomes in Catalan for those measures evaluating essentially linguistic knowledge or use, i.e., ungrammatical but semantically sound sentences in the SJT and the microstructural narrative analysis. Our findings, however, revealed that bilingual speakers showed comparable metalinguistic, grammatical knowledge and narrative abilities at both the macrostructural and microstructural level in both Spanish and Catalan. As far as we know, this is the first study that has investigated the metalinguistic and narrative abilities of individuals with ID in both of their languages and also the first study that has made evident not only that early sequential bilinguals with genetic syndromes can perform comparably to monolingual speakers with the same syndrome in their non-dominant language but also that they can show comparable metalinguistic and narrative abilities in both of their languages regardless of the language dominance factor.

8.2. The effects of bilingualism on the Prader-Willi Syndrome population

A comparative analysis between the effects of bilingualism in the three areas investigated in this dissertation reveal, overall, parallel results: no significant differences between monolingual and bilingual speakers with PWS, with, as already mentioned, the exception of the lexical variability measure in spontaneous data production, where bilinguals outperformed monolinguals. These findings are unequivocal and align with previous research not revealing a negative effect of bilingualism for non-TD individuals (see Kay-Raining Bird et al., 2016 for an overview). However, the discussion that remains opens is how to interpret our results in the light of the BAH, whether as a failure to detect it or as a sign of its inexistence (see Valian, 2015 for a complete discussion about these two positions).

In the case of the TD individuals, this debate is not new and each position has followers and detractors.[75] For instance, Bialystok and collaborators are well-known advocates of the so-called BA (Bialystok et al., 2008; Kroll & Bialystok, 2013; among others) while Paap and colleagues (Paap, 2014; Paap et al., 2015) represent what can be called the critical position (see Bialystok, 2016 for a comprehensive review of the ongoing debate). Going back in history, it has been the case that since Peal and Lambert's (1962) results showing positive effects of bilingualism at the cognitive level, different researchers have devoted their research to study the effects of bilingualism on dual speakers' cognition, with the EC domain being one of the areas that has attracted more attention and which has intensified during the last years. At the beginning, results were predominantly positive, which led to the BAH proposal. However, as the field grew so did the variability of the results and the skepticism about the existence of such BA (Bialystok, 2016). In fact, it has been recently suggested that the BA could be the result of a publication bias, where those studies supporting the BA would have had more chances of being published than those challenging it (de Bruin et al., 2015).

Bialystok (2016) argues that in order to have an accurate picture of the real impact of bilingualism, an in-depth analysis is needed with particular focus on different aspects, among which a proper data interpretation stands out. Bialystok (2016) discusses that one of the main

[75] As already mentioned in the introduction, although the more vivid debate principally lays at the EC domain, the arguments presented within this subfield has been extended here to the other two areas under study, for which a BA has also been proposed (see chapter 2).

reasons that has led some researchers to challenge the existence of a BA is the fact of not finding differences between monolinguals and bilinguals in the different studies conducted. However, as she points out, it is important to have in mind that not finding a BA does not necessary imply its nonexistence, since what null effects are telling us is that we cannot reject the null hypothesis but under no circumstances can they be taken as an irrefutable proof of no differences between groups (Bialystok, 2016, p. 528). Another important question which Bialystok emphasizes when evaluating results is the importance of taking into account the population included in the study, since, as already discussed, the young adult population has been repetitively proven to be a particularly challenging one for detecting a potential BA (see Bialystok, 2016, or Vinerte, 2018, for overviews), especially when the experimental tasks administered are simple and not challenging (e.g., Costa et al., 2009). Two main possible explanations have been proposed for this effect: on the one hand, the argument that the young adult population would be at their top processing level (Kroll & Bialystok, 2013) and, on the other hand, the argument suggesting that adolescents and young adults (also extensible to children) would be in a vital phase where they would be exposed to different "cognitively challenging activities", in Valian's words, that could result in similar effects to those proposed for bilingualism (Valian, 2015, p. 19). Nevertheless, it is a fact that the results are not clear-cut and while there are studies centered on the young adult population that have revealed a BA (e.g., Bialystok et al., 2008), others have shown a null effect of bilingualism in populations where traditionally a BA has been shown to be more robust, as in the case of children (e.g., Duñabeitia et al., 2014). These divergences suggest that even though the developmental stage may play a role on the results obtained, other factors may condition the appearance or not of a BA. Luk and Bialystok (2013) suggest that bilingualism should not be reduced to a mere yes/no classification but that it would be better described as a continuum involving different types (or degrees) of bilinguals. Hence, it is very important to carefully take into consideration the bilinguals' characteristics when designing a study and when interpreting the results (Bialystok, 2018).

In this dissertation, given the difficulty to recruit participants with PWS, especially bilingual speakers, the participant sample included a wide age range, although both groups were comparable in this respect, as the majority were teenagers or young adults. This, together with the fact that the descriptive data at the EC, metalinguistic and narrative macrostructural level suggested better outcomes for bilinguals even though they were compared to monolinguals in their non-dominant language, makes us adhere to Kroll and Bialystok (2013) and Valian's (2015) interpretation, among many other researchers, which highlights the difficulty to isolate a BA under certain conditions. In this case the small sample, the participants' developmental stage and the language dominance factor might be hindering a potential BA. For the moment, our results do not allow us to make inferences beyond pointing to an overall effect of bilingualism among adolescents and young adults with PWS akin to that found in analogous TD populations using similar tasks to the ones used in this dissertation.

It is premature to draw conclusions about the existence or inexistence of a BA among the PWS population and more research is needed to determine whether the relationship between bilingualism and EC, metalinguistic, and narrative abilities previously described for the TD population holds for individuals with ID, as some recent studies seem to suggest for DLD-SLI and ASD individuals (Baldimtsi et al., 2020; Gonzalez-Barrero & Nadig, 2017, 2019a; Peristeri et al., 2020; among others), or whether these individuals show an idiosyncratic behavior.

8.3. Contributions, implications, limitations and future research

8.3.1. Contributions

To the best of our knowledge, beyond two previous case studies (García-Alcaraz, 2018; Liceras & García-Alcaraz, 2019), this is the first piece of research that has specifically addressed the study of bilingualism in the PWS population with a group study design. At the theoretical level, our results make an important contribution to the field of bilingualism in general and to the subfield of bilingualism in atypical language development in particular. On the one hand, this study provides valuable data about the nonverbal IQ, receptive vocabulary and sentence recall abilities of the PWS population, and, on the other hand, it offers novel data about the effects of bilingualism on the EC, metalinguistic, and narrative abilities of these individuals. Despite bilingualism being the focus of our research, the findings of this dissertation are not only pioneering in providing insights on how bilingualism may shape the cognitive and linguistic abilities of bilinguals with PWS but also on how to complement previous research intended to acquire more knowledge about the cognitive and linguistic profile of monolinguals with PWS since, as stated in chapter 3, the information available for PWS is limited when compared to that available for other genetic syndromes (Lewis, 2006). As extensively discussed in these pages, the study of bilingualism on individuals with DD is scarce and, even if growing, it is still in its first stages, being especially limited in the case of individuals with ID, which is nearly restricted to DS (see Kay-Raining Bird et al., 2016 for a review). Thus, for the first time, at least in the terms presented here, the effects of bilingualism in a non-TD population with ID have been scrutinized at both the cognitive and the linguistic levels using both experimental (offline and online) and oral spontaneous data.

At the experimental level, as already mentioned, our findings contribute to this line of research by providing new data from a different genetic syndrome (PWS), an understudied linguistic combination (Spanish-Catalan), an understudied bilingual profile (early sequential bilinguals) and an unexplored developmental stage (mainly adolescents and adults). Another contribution of this dissertation at the experimental level is the different experimental tasks created to address the different objectives set. The EC and metalinguistic tasks used in this dissertation are not novel, since they are established tasks within the field of bilingualism to assess the participants' EC and metalinguistic abilities. However, they are a contribution in the sense that they have been adapted for being used for the first time in both Spanish and Catalan, paying close attention to the target population and their idiosyncratic needs. The experimental tasks were designed in view of the nature of the items, the number of items per condition, and the task procedure, since these factors were crucial to maintain the participants' attention and to ensure data reliability. Likewise, in an attempt to get the best outcomes from our results, we adopted different statistical approaches taking into account the nature of the data obtained for each experimental task and for each specific research question. Between-group analyses including repeated measures were run with mixed-effects models analyses, which are powerful state-of-the-art statistical analyses that take into account all data points and account for by-participant and by-item variability. This type of analysis is gaining force within the field and this, together with all the aforementioned contributions, makes this dissertation unique.

8.3.2. Implications

The results obtained within the framework of this dissertation support the claims, also made in previous research, that individuals with DD do become bilingual, within the intrinsic limitations of their condition, and, that speaking two languages does not show signs of negative effects (Kay-Raining Bird et al., 2016; among others). As Kay-Raining Bird et al. (2016) maintain, the results derived from this type of research are specially relevant to help parents with children with DD, as well as to practitioners working with this population, to make informed decisions about the possibility of these individuals becoming bilingual. In this sense, the singularity of the Catalan context (high presence of both languages within all social and media spheres) and the proximity between these two languages, offers a unique scenario to gain a better understanding of the effects of bilingualism in these individuals.

Unlike what may be the case in other bilingual contexts, these days Catalan non-TD individuals, at least those with mild to moderate ID, become bilingual "by default" rather than as a result of a conscious parental decision. For the majority of our bilingual participants Catalan was the family main language of communication and the language of education. However, despite this, their command of Spanish did not differ much from that of their monolingual peers, neither at the experimental level, whose results have been presented and discussed here, nor at a more personal level (i.e., the researcher's perception when interacting with them during data collection). Both Catalan and Spanish are intrinsically present in the Catalan society and, nowadays, irrespective of each individual's dominant language, all are exposed to the other language via instruction at school or naturalistically, be it due to contact with friends and/or family members, leisure activities or access to the media and social media platforms.

This dissertation compared monolinguals with bilinguals mainly in the bilinguals' non-dominant language and our findings not only showed that both groups performed comparably but that bilinguals even outperformed Spanish monolinguals in some aspects, as was the case with the lexical variability measure in the narrative task. Thus, what we know so far is that bilinguals with genetic syndromes are not only comparable to their monolingual peers in their command of the dominant language (Feltmate & Kay-Raining Bird, 2008; Kay-Raining Bird et al., 2005; among others) but also in their command of the non-dominant language, or at least with respect to some cognitive and linguistic aspects, even if they are early-sequential bilinguals. These findings seem to distance themselves from Kay-Raining Bird et al.'s (2016) proposal suggesting that in order to match their monolingual peers, sequential bilinguals with DD could need more time. Our two groups, despite not being homogeneous in terms of age at the within-group level, as we have discussed and justified, did not significantly differ at the between-group level. Nonetheless, in order to have a better understanding of the language developmental process in different types and degrees of bilingualism, an exhaustive contrastive longitudinal study is needed. On the other hand, our findings align with Kay-Raining Bird et al.'s (2016) claim about the importance of the amount of language exposure for language development in general and for the development of the non-dominant language in particular. In our case, language exposure was key and definitely the proximity between both languages might have also played a decisive a role. Be that as it may, what our results have made evident is that when both factors converge (i.e., high exposure and language proximity) the outcomes seem to be quite encouraging.

It is premature to draw educational or pedagogical implications from these results and more research is needed to be in position to formulate an intervention plan for the classroom. Knowing more about this population's strengths and shortcomings in relation to the TD population would provide educational professionals and practitioners with valuable information to offer this population a more tailored curriculum that could result in better academic outcomes. For example, a pedagogical implication that may derive from this type of research and that can be useful to illustrate our point may be the support needed by these individuals with narrative production. It has been repetitively shown that these individuals, be they monolingual or bilingual, have problems when narrating, a very common but complex activity that has implications for their social relationships (Lewis, 2006). Thus, intervention and educational professionals should learn from empirical research highlighting these deficiencies and explicitly address them, or, if they are already doing so, they could increment the time dedicated to working on narrative production and reconsidering the way in which they are addressing this issue. This implication is not new and has been previously discussed in Garayzábal-Heinze et al. (2012) and Lewis et al. (2002) for monolingual speakers with PWS and in García-Alcaraz (2018) for a bilingual speaker with PWS. However, it is worthwhile to revisit it here, since we have more data from a group study design comparing monolingual and bilingual individuals. This makes us aware of the importance of conducting research with these populations in order to encourage transferring this knowledge to the classroom. Among the limited linguistic research available for the PWS population, narratives have been an area of relative interest and this has made it possible to make a first attempt of discussing a pedagogical implication. We are confident that in the near future more implications would follow in other domains when more research is conducted, and more findings become available.

As Paradis (2016, p. 80) points out "our knowledge of bilingualism and the capacity for it in children with DD has only scratched the surface", thus we join her and many other researchers working on the field (Cleave et al., 2014; Feltmate & Kay-Raining Bird, 2008; Kay-Raining Bird et al., 2005; among others) in advocating for the encouragement, whenever possible, of bilingualism among individuals with DD unless it is proven that bilingualism has significant negative effects beyond those already described for the TD population, something that seems highly unlikely if we take into account the consistency of the non-negative results so far.

8.3.3. Limitations and future research

Although the number of participants is considerable taking into account the population included in the study, the fact of not counting with a larger sample similar to those included in studies exploring the effects of bilingualism among TD individuals could be perceived as a limitation. Taking this into account, the conclusions drawn here need to be taken cautiously, since, for the moment they are circumscribed to this participants' sample and we need more data to extrapolate our findings to the PWS population at large. Similarly, more data is needed in order to be able to perform more powerful statistical analyses. This dissertation incorporated a fine-grained mixed-effects models analysis to have a better understanding of how monolinguals and bilinguals with PWS compared in the different experimental tasks including repeated measures. However, due to the data available, we needed to be particularly careful not to overfit the models in order to ensure convergence and, as a result, we broke the analyses into two phases.[76] Predictably, expanding the participants' sample size will not only allow us to run the different analyses including all the predictors at once but also to investigate interactions between variables unexplored so far, such as, for instance, the interaction between the three standardized test scores and the different experimental conditions. Having more data would also increase the chances of obtaining convergence with more complex models, especially at the random effects structure level. All of this would provide a more precise and richer picture of the results obtained.

Future research centered on elucidating the effects of bilingualism in non-TD individuals with special emphasis on testing the BAH should try to administer different cognitive and linguistic tasks to two more homogenous groups in terms of age. It would also be advisable to avoid the young adult population and to recruit either children or older adults, since these two age

[76] Different to the mixed-effects model analyses, regression analyses were run with all the predictors at once.

groups have been said to be the ones for whom it would be more feasible to detect a potential beneficial effect of bilingualism. Also, it would be noteworthy to study the role of language dominance, since it could be the case that a potential BA could arise when comparing monolinguals and bilinguals with PWS in the bilinguals' dominant language. For instance, in the specific case of Spanish-Catalan bilingualism, it would be interesting to see how Spanish dominant and Catalan dominant bilinguals compare between them and in relation to the Spanish monolingual group with respect to different cognitive and linguistic abilities.

Especially relevant would also be the exploration of how monolinguals and bilinguals with different PWS subtypes compare at the cognitive and linguistic levels, since previous research has pointed to possible differences between the two PWS subtypes with a higher incidence: DEL and UPD (see chapter 3). We were not able to do so because more than 70% of our sample were individuals belonging to the DEL subtype. To address this comparison, however important, would be challenging, since individuals with UPD represent a 20-30% of the PWS population and individuals with imprinting only a 1-3% (Cassidy et al., 2012; Foundation for Prader-Willi Research, n.d.). More recently, it has been suggested that PWS individuals that have received growth hormone treatment show better full and verbal IQ scores than those who did not and that the effects were larger when the treatment started during the first year of age (Dykens et al., 2017). The effects of how the growth hormone treatment, if received, influenced the results obtained was not explored here, as it was not available in the previous literature reviewed, but it is definitely something that should be carefully looked at in future studies.

In relation to the experimental task design, the WLJT proposed did not prove to be a very enlightening task to assess the participants' metalinguistic abilities because both monolinguals and bilinguals performed at ceiling. This indicates that this task was not challenging enough to detect the participants' underlying metalinguistic abilities. This is valuable information to take into account in future research when designing new tasks geared towards this population. Nevertheless, it is worth noting that, as discussed in chapter 6, from an observational point of view, participants took longer and produced more dubitative answers for incongruent items. Therefore, in order to (dis)prove this observation, the next step would be to adapt this task to an online version to measure participants' response latency. Also, it would be interesting to adapt the SJT to an online task, as in Bialystok et al. (2014). This way we would have a more comprehensive view of how these individuals process grammatical meaningless sentences (metalinguistic measure) and ungrammatical meaningful sentences (linguistic measure). Future research should also explicitly focus on how monolingual and bilingual speakers with PWS deal with different morphosyntactic language phenomena, since a better understanding of these individuals' linguistic profile is needed in order to carry out intervention policies intended to offer them with an educational curriculum tailored to their needs.

In this study we neither included a TD monolingual nor a TD bilingual group because our main objective was to specifically study the effects of bilingualism on the PWS population. However, future research should explore whether and how individuals with PWS differ from the TD population when performing different cognitive and linguistic tasks. This would allow us to detect relative strengths and shortcomings that, properly identified, could potentially translate into better intervention measures at the academic level. Another issue especially relevant would be to investigate how bilingual individuals with different genetic syndromes compare when performing different cognitive and linguistic tasks. In this respect, some between-syndrome comparison studies have been carried out with monolingual individuals. For instance, Garayzábal-Heinze et al. (2012) compared the narrative abilities of individuals with PWS, WS and Smith-Magenis

Syndrome, Díez-Itza et al. (2019), on their part, contrasted the morphological abilities of individuals with DS and WS and, Woodcock et al. (2009) compared the EC abilities of individuals with PWS and Fragile X syndrome, to name a few examples. It is obvious that more research is needed, especially with bilinguals, since we need to determine how individuals with different genetic syndromes resemble or differ, so that we can have a clearer picture of whether some cognitive and linguistic characteristics and the potential effects of bilingualism are common across genetic disorders or are rather syndrome specific. Again, having this information would translate into better intervention practices and would help intervention professionals to address general and specific needs when dealing with a heterogeneous group of individuals with these characteristics.

We would like to conclude by acknowledging that, as with every piece of research, this dissertation presents limitations which, nonetheless, do not undermine the relevance of the study, since these results represent a first attempt to better understand the effects of bilingualism in the PWS population and constitute a basis for continuing to build new knowledge in this rather virgin research area. Different researchers have devoted an important body of research to raise awareness about the importance of studying bilingualism among individuals with ID and have made this field take off. With this dissertation, we hope to contribute to taking the field one step forward and it is our hope that this will be our first group study, of many others to come, investigating the effects of bilingualism in individuals with genetic syndromes. We also hope that other researchers in the field, as well as newcomers, will join with new data to make the field grow and advance.

References

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernández, M., Scifo, P., Keim, R., Cappa, S. F., & Costa, A. (2012). Bilingualism Tunes the Anterior Cingulate Cortex for Conflict Monitoring. *Cerebral Cortex*, *22*(9), 2076–2086. https://doi.org/10.1093/cercor/bhr287

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A Systematic Review and Meta-Analysis of the Cognitive Correlates of Bilingualism. *Review of Educational Research*, *80*(2), 207–245. https://doi.org/10.3102/0034654310368803

Åkefeldt, A., Åkefeldt, B., & Gillberg, C. (1997). Voice, speech and language characteristics of children with Prader-Willi syndrome. *Journal of Intellectual Disability Research*, *41*(4), 302–311. https://doi.org/10.1111/j.1365-2788.1997.tb00713.x

Alexander, R. C., Van Dyke, D. C., & Hanson, J. W. (1995). Overview. In L. R. Greenswag & R. C. Alexander (Eds.), *Management of Prader-Willi Syndrome* (2nd ed., pp. 3–17). Springer-Verlag.

Alloway, T. P., & Ledwon, F. (2014). Working memory and sentence recall in children. *International Journal of Educational Research*, *65*, 1–8. https://doi.org/10.1016/j.ijer.2013.12.001

Altman, C., Armon-Lotem, S., Fichman, S., & Walters, J. (2016). Macrostructure, microstructure, and mental state terms in the narratives of English-Hebrew bilingual preschool children with and without specific language impairment. *Applied Psycholinguistics*, *37*(1), 165–193. https://doi.org/10.1017/S0142716415000466

Alves-Soares, L. (2020). *Investigating the Portuguese-English Bilingual Mental Lexicon: Effects of Crosslinguistic Orthographic and Phonological/Phonetic Overlap* [Doctoral Dissertation, University of Ottawa]. https://ruor.uottawa.ca/handle/10393/41153

American Association on Intellectual and Developmental Disabilities. (n.d.). *Frequently Asked Questions on Intellectual Disability*. Retrieved September 4, 2020, from https://www.aaidd.org/intellectual-disability/definition/faqs-on-intellectual-disability

American Association on Intellectual and Developmental Disabilities. (2020). *Definition of Intellectual Disability*. https://www.aaidd.org/intellectual-disability/definition

American Psychiatric Association. (2018). *What Is Autism Spectrum Disorder?* https://www.psychiatry.org/patients-families/autism/what-is-autism-spectrum-disorder#:~:text=Autism spectrum disorder (ASD) is,are different in each person.

American Psychiatric Association. (2020). *What is Intellectual Disability?* https://www.psychiatry.org/patients-families/intellectual-disability/what-is-intellectual-disability

Andreou, M. (2015). *The effects of bilingualism on verbal and non verbal cognition: The micro- and macro-structure of narratives in the weak and the dominant language of the bilingual child* [Doctoral dissertation, Aristotle University of Thessaloniki]. https://www.didaktorika.gr/eadd/handle/10442/36739?locale=en

Antón, E., Duñabeitia, J. A., Estévez, A., Hernández, J. A., Castillo, A., Fuentes, L. J., Davidson, D. J., & Carreiras, M. (2014). Is there a bilingual advantage in the ANT task? Evidence from children. *Frontiers in Psychology*, *5*, 398. https://doi.org/10.3389/fpsyg.2014.00398

Ardila, A. (2012). Ventajas y desventajas del bilingüismo. *Forma y Función*, *25*(2), 99–114.

Astheimer, L., Janus, M., Moreno, S., & Bialystok, E. (2014). Electrophysiological measures of attention during speech perception predict metalinguistic skills in children. *Developmental Cognitive Neuroscience*, *7*, 1–12. https://doi.org/10.1016/j.dcn.2013.10.005

Atkin, K., & Lorch, M. P. (2007). Language development in a 3-year-old boy with Prader-Willi syndrome. *Clinical Linguistics and Phonetics*, *21*(4), 261–276. https://doi.org/10.1080/02699200701243865

Baldimtsi, E., Peristeri, E., Tsimpli, I. M., & Durrleman, S. (2020). The Impact of Bilingualism on Theory of Mind and Executive Functions in TD and ASD. In M. M. Brown & A. Kohut (Eds.), *Proceedings of the 44th Boston University Conference on Language Development* (pp. 40–52). Cascadilla Press.

Baldimtsi, E., Peristeri, E., Tsimpli, I. M., & Nicolopoulou, A. (2016). Bilingual Children with High Functioning Autism Spectrum Disorder: Evidence from Oral Narratives and Non-verbal Executive Function Tasks. In J. Scott & D. Waughtal (Eds.), *Proceedings of the 40th annual Boston University Conference on Language Development* (pp. 18–31). Cascadilla Press.

Barisnikov, K., Van Der Linden, M., & Poncelet, M. (1996). Acquisition of new words and phonological working memory in Williams syndrome: A case study. *Neurocase*, *2*(5), 395–404. https://doi.org/10.1093/neucas/2.5.395-a

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barton, K. (2020). *MuMIn: Multi-Model Inference. R package version 1.43.17*. https://cran.r-project.org/package=MuMIn

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bell, N. L., Lassiter, K. S., Matthews, T. D., & Hutchinson, M. B. (2001). Comparison of the Peabody Picture Vocabulary Test - Third edition and Wechsler Adult Intelligence Scale - Third edition with University Students. *Journal of Clinical Psychology*, *57*(3), 417–422. https://doi.org/10.1002/jclp.1024

Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Psychology Press.

Bertella, L., Girelli, L., Grugni, G., Marchi, S., Molinari, E., & Semenza, C. (2005). Mathematical skills in Prader – Willi Syndrome. *Journal of Intellectual Disability Research*, *49*(2), 159–169. https://doi.org/10.1111/j.1365-2788.2004.00634.x

Bialystok, E. (1986a). Children's Concept of Word. *Journal of Psycholinguistic Research*, *15*(1), 13–32. https://doi.org/10.1007/BF01067389

Bialystok, E. (1986b). Factors in the Growth of Linguistic Awareness. *Child Development*, *57*(2), 498–510. https://doi.org/10.2307/1130604

Bialystok, E. (1987). Influences of bilingualism on metalinguistic development. *Interlanguage Studies Bulletin (Utrecht)*, *3*(2), 154–166. https://doi.org/10.1177/026765838700300205

Bialystok, E. (1988). Levels of Bilingualism and Levels of Linguistic Awareness. *Developmental Psychology*, *24*(4), 560–567. https://doi.org/10.1037/0012-1649.24.4.560

Bialystok, E. (1992). Attentional control in children's metalinguistic performance and measures of field independence. *Developmental Psychology*, *28*(4), 654–664. https://doi.org/10.1037/0012-1649.28.4.654

Bialystok, E. (2001). *Bilingualism in development: Language, Literacy, and Cognition.* Cambridge University Press.

Bialystok, E. (2016). The signal and the noise: Finding the pattern in human behavior. *Linguistic Approaches to Bilingualism*, *6*(5), 517–534. https://doi.org/10.1075/lab.15040.bia

Bialystok, E. (2018). Bilingualism and executive function: What's the connection? In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual Cognition and Language: The state of the science across its subfields* (pp. 283–305). John Benjamins. https://doi.org/https://doi.org/10.1075/sibil.54.13bia

Bialystok, E., & Barac, R. (2012). Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition*, *122*(1), 67–73. https://doi.org/10.1016/j.cognition.2011.08.003

Bialystok, E., Barac, R., Blaye, A., & Poulin-Dubois, D. (2010). Word Mapping and Executive Functioning in Young Monolingual and Bilingual Children. *Journal of Cognition and Development*, *11*(4), 485–508. https://doi.org/https://doi.org/10.1080/15248372.2010.516420

Bialystok, E., Craik, F. I. M., & Luk, G. (2008). Cognitive Control and Lexical Access in Younger and Older Bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 859–873. https://doi.org/10.1037/0278-7393.34.4.859

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, *16*(4), 240–250. https://doi.org/10.1016/j.tics.2012.03.001

Bialystok, E., & Luk, G. (2012). Receptive vocabulary differences in monolingual and bilingual adults. *Bilingualism: Language and Cognition*, *15*(2), 397–401. https://doi.org/10.1017/S136672891100040X

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). 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 Psycholinguistics*, *19*(1), 69–85. https://doi.org/10.1017/s0142716400010584

Bialystok, E., & Mitterer, J. (1987). Metalinguistic Differences Among Three Kinds of Readers. *Journal of Educational Psychology*, *79*(2), 147–153. https://doi.org/10.1037/0022-0663.79.2.147

Bialystok, E., Peets, K. F., & Moreno, S. (2014). Producing bilinguals through immersion education: Development of metalinguistic awareness. *Applied Psycholinguistics*, *35*(1), 177–191. https://doi.org/10.1017/S0142716412000288

Bialystok, E., Poarch, G. J., Luo, L., & Craik, F. I. M. (2014). Effects of Bilingualism and Aging on Executive Function and Working Memory. *Psychology and Aging*, *29*(3), 696–705. https://doi.org/10.1037/a0037254

Bialystok, E., & Ryan, E. B. (1985). Toward a Definition of Metalinguistic Skill. *Merrill-Palmer Quarterly*, *31*(3), 229–251.

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE-2, C. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *The Journal of Child Psychology and Psychiatry*, *58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Bitetti, D., Hammer, C. S., & López, L. M. (2020). The narrative macrostructure production of Spanish-English bilingual preschoolers: Within- and cross-language relations. *Applied Psycholinguistics*, *41*(1), 79–106. https://doi.org/10.1017/S0142716419000419

Boerma, T., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2016). Narrative abilities of monolingual and bilingual children with and without language impairment: implications for clinical practice. *International Journal of Language and Communication Disorders*, *51*(6), 626–638. https://doi.org/10.1111/1460-6984.12234

Boix-Fuster, E., & Sanz, C. (2008). Language and identity in Catalonia. In M. Niño-Murcia & J. Rothman (Eds.), *Bilingualism and identity: Spanish at the crossroads with other languages* (pp. 87–106). John Benjamins. https://doi.org/10.1075/sibil.37

Bolker, B. (2014). *Converge error for development version of lme4*. https://stackoverflow.com/questions/21344555/convergence-error-for-development-version-of-lme4

Bowers, D. L. (2012). *Improving Literacy Through Instruction and Community Experiences: The effects of a summer program on literacy outcomes of students from low SES homes.* [Doctoral dissertation, University of Tennessee]. https://dc.uthsc.edu/dissertations/29/

Branson, C. (1981). Speech and language characteristics of children with Prader-Willi syndrome. In V. A. Holm, S. Sulzbacher, & P. L. Pipes (Eds.), *The Prader-Willi syndrome* (pp. 179–183). University Park Press.

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (1990). *Test of Nonverbal Intelligence (TONI-2)*. Pro-ed.

Brown, R. (1973). Development of the First Language in the Human Species. *American Psychologist*, *28*(2), 97–106. https://doi.org/10.1037/h0034209

Brydges, C. R., Anderson, M., Reid, C. L., & Fox, A. M. (2013). Maturation of Cognitive Control: Delineating Response Inhibition and Interference Suppression. *PLoS ONE*, *8*(7), e69826. https://doi.org/10.1371/journal.pone.0069826

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition*, *1*(1), 1–20. https://doi.org/10.5334/joc.10

Bueno Díez, M., & Caixàs Pedragós, A. (2014). Prader-Willi syndrome and hyperphagia: A challenge to investigate. *Endocrinología y Nutrición*, *61*(3), 121–122. https://doi.org/10.1016/j.endoen.2014.02.016

Buiting, K., Färber, C., Kroisel, P., Wagner, K., Brueton, L., Robertson, M. E., Lich, C., & Horsthemke, B. (2000). Imprinting centre deletions in two PWS families: Implications for diagnostic testing and genetic counselling. *Clinical Genetics*, *58*(4), 284–290. https://doi.org/10.1034/j.1399-0004.2000.580406.x

Burgoyne, K., Duff, F. J., Nielsen, D., Ulicheva, A., & Snowling, M. J. (2016). Bilingualism and Biliteracy in Down Syndrome: Insights From a Case Study. *Language Learning*, *66*(4), 945–971. https://doi.org/10.1111/lang.12179

Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach (2nd ed.)*. Springer. https://doi.org/10.2307/3802723

Butler, M. G., Lee, P. D. K., & Whitman, B. Y. (Eds.). (2006). *Management of Prader-Willi Syndrome* (Third Edit). Springer New York. https://doi.org/10.1007/978-0-387-33536-0

Byers-Heinlein, K., & Lew-Williams, C. (2013). Bilingualism in the early years: What the Science Says. *LEARNing Landscapes*, *7*(1), 95–112. https://doi.org/10.36510/learnland.v7i1.632

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

Carson, R. (2018). *Processing Grammatical and Notional Number Information in English and French* [Doctoral dissertation, University of Ottawa]. https://ruor.uottawa.ca/handle/10393/38309

Cassidy, S. B. (1997). Prader-Willi syndrome. *Journal of Medical Genetics*, *34*(11), 917–923. https://doi.org/10.1136/jmg.34.11.917

Cassidy, S. B., & Driscoll, D. J. (2009). Prader-Willi syndrome. *European Journal of Human Genetics*, *17*(1), 3–13. https://doi.org/10.1038/ejhg.2008.165

Cassidy, S. B., Schwartz, S., Miller, J. L., & Driscoll, D. J. (2012). Prader-Willi syndrome. *Genetics in Medicine*, *14*(1), 10–26. https://doi.org/10.1038/gim.0b013e31822bead0

Chen, L., & Yan, R. (2011). Development and use of English evaluative expressions in narratives of Chinese-English bilinguals. *Bilingualism: Language and Cognition*, *14*(4), 570–578. https://doi.org/10.1017/S1366728910000362

Chevalère, J., Postal, V., Jauregui, J., Copet, P., Laurier, V., & Thuilleaux, D. (2013). Assessment of Executive Functions in Prader-Willi Syndrome and Relationship with Intellectual Level. *Journal of Applied Research in Intellectual Disabilities*, *26*(4), 309–318. https://doi.org/10.1111/jar.12044

Chevalère, J., Postal, V., Jauregui, J., Copet, P., Laurier, V., & Thuilleaux, D. (2015). Executive functions and Prader-Willi Syndrome: Global Deficit Linked with Intellectual Level and Syndrome-Specific Associations. *American Journal on Intellectual and Developmental Disabilities*, *120*(3), 215–229. https://doi.org/10.1352/1944-7558-120.3.215

Chung-Fat-Yim, A., Himel, C., & Bialystok, E. (2019). The impact of bilingualism on executive function in adolescents. *International Journal of Bilingualism*, *23*(6), 1278–1290. https://doi.org/10.1177/1367006918781059

Cimolin, V., Galli, M., Grugni, G., Vismara, L., Albertini, G., Rigoldi, C., & Capodaglio, P. (2010). Gait patterns in Prader-Willi and Down syndrome patients. *Journal of NeuroEngineering and Rehabilitation*, *7*(1), 1–8. https://doi.org/10.1186/1743-0003-7-28

Cleave, P. L., Girolametto, L., Chen, X., & Johnson, C. J. (2010). Narrative abilities in monolingual and dual language learning children with specific language impairment. *Journal of Communication Disorders*, *43*(6), 511–522. https://doi.org/10.1016/j.jcomdis.2010.05.005

Cleave, P. L., Kay-Raining Bird, E., Trudeau, N., & Sutton, A. (2014). Syntactic bootstrapping in children with Down syndrome: The impact of bilingualism. *Journal of Communication Disorders*, *49*, 42–54. https://doi.org/10.1016/j.jcomdis.2014.02.006

Coderre, E. L., & Van Heuven, W. J. B. (2014). Electrophysiological Explorations of the Bilingual Advantage: Evidence from a Stroop Task. *PLOS ONE*, *9*(7), e103424. https://doi.org/10.1371/journal.pone.0103424

Coderre, E. L., Van Heuven, W. J. B., & Conklin, K. (2013). The timing and magnitude of Stroop interference and facilitation in monolinguals and bilinguals. *Bilingualism: Language and Cognition*, *16*(2), 420–441. https://doi.org/10.1017/S1366728912000405

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge. https://doi.org/10.4324/9780203771587

Conners, F. A., Rosenquist, C. J., Atwell, J. A., & Klinger, L. G. (2000). Cognitive Strengths and Weaknesses Associated with Prader-Willi Syndrome. *Education and Training in Mental Retardation and Developmental Disabilities*, *35*(4), 441–448.

Cook, V. (1997). The consequences of bilingualism for cognitive processing. In A. M. B. de Groot & J. Kroll (Eds.), *Tutorials in Bilingualism: Psycholinguistic Perspectives* (pp. 279–299). Psychology Press.

Copet, P., Jauregui, J., Laurier, V., Ehlinger, V., Arnaud, C., Cobo, A. M., Molinas, C., Tauber, M., & Thuilleaux, D. (2010). Cognitive profile in a large french cohort of adults with Prader-Willi syndrome: Differences between genotypes. *Journal of Intellectual Disability Research*, *54*(3), 204–215. https://doi.org/10.1111/j.1365-2788.2010.01251.x

Costa, A., Albareda, B., & Santesteban, M. (2008). Assessing the presence of lexical competition across languages: Evidence from the Stroop task. *Bilingualism: Language and Cognition*, *11*(1), 121–131. https://doi.org/10.1017/S1366728907003252

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

Cromdal, J. (1999). Childhood bilingualism and metalinguistic skills: Analysis and control in young Swedish-English bilinguals. *Applied Psycholinguistics*, *20*(1), 1–20. https://doi.org/10.1017/s0142716499001010

Cummins, J. (1976). The Influence of Bilingualism on Cognitive Growth: A Synthesis of Research Findings and Explanatory Hypotheses. *Working Papers On Bilingualism*, *9*, 1–44.

Curfs, L. M., Wiegers, A. M., Sommers, J. R. M., Borghgraef, M., & Fryns, J. P. (1991). Strengths and weaknesses in the cognitive profile of youngsters with Prader-Willi syndrome. *Clinical Genetics*, *40*(6), 430–434. https://doi.org/10.1111/j.1399-0004.1991.tb03114.x

Davidson, D. J., Zacks, R. T., & Williams, C. C. (2003). Stroop interference, Practice, and Aging. *Aging, Neuropsychology, and Cognition*, *10*(2), 85–98. https://doi.org/10.1076/anec.10.2.85.14463

de Bruin, A., Samuel, A. G., & Duñabeitia, J. A. (2018). Voluntary language switching: When and why do bilinguals switch between their languages? *Journal of Memory and Language*, *103*, 28–43. https://doi.org/10.1016/j.jml.2018.07.005

de Bruin, A., Samuel, A. G., & Duñabeitia, J. A. (2020). Examining bilingual language switching across the lifespan in cued and voluntary switching contexts. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(8), 759–788. https://doi.org/10.1037/xhp0000746

de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive Advantage in Bilingualism: An Example of Publication Bias? *Psychological Science*, *26*(1), 99–107. https://doi.org/10.1177/0956797614557866

Defloor, T., Van Borsel, J., & Curfs, L. M. (2000). Speech Fluency in Prader-Willi Syndrome. *Journal of Fluency Disorders*, *25*(2), 85–98. https://doi.org/10.1016/S0094-730X(00)00026-7

Defloor, T., Van Borsel, J., Curfs, L. M., & De Bodt, M. (2001). Aerodynamic and acoustic characteristics of voice in Prader-Willi syndrome. *Journal of Voice*, *15*(2), 284–290. https://doi.org/10.1016/S0892-1997(01)00029-7

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

Díez-Itza, E., Martínez, V., Pérez, V., & Fernández-Urquiza, M. (2018). Explicit Oral Narrative Intervention for Students with Williams Syndrome. *Frontiers in Psychology*, 2337. https://doi.org/10.3389/fpsyg.2017.02337

Díez-Itza, E., Miranda, M., Pérez, V., & Martínez, V. (2019). Profiles of grammatical morphology in Spanish speaking adolescents with Williams Syndrome and Down Syndrome. In E. Aguilar-Mediavilla, L. Buil-Legaz, R. López-Penadés, V. A. Sanchez-Azanza, & D. Adrover-Roig (Eds.), *Atypical Language Development in Romance Languages* (pp. 219–234). John Benjamins. https://doi.org/10.1075/z.223.13die

Dimitropoulos, A., Ferranti, A., & Lemler, M. (2013). Expressive and receptive language in Prader-Willi syndrome: Report on genetic subtype differences. *Journal of Communication Disorders*, *46*(2), 193–201. https://doi.org/10.1016/j.jcomdis.2012.12.001

Dimitropoulos, A., Ho, A., & Feldman, B. (2013). Social Responsiveness and Competence in Prader-Willi Syndrome: Direct Comparison to Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *43*(1), 103–113. https://doi.org/10.1007/s10803-012-1547-3

Dimitropoulos, A., & Schultz, R. T. (2007). Autistic-like symptomatology in Prader-Willi syndrome: A review of recent findings. *Current Psychiatry Reports*, *9*(2), 159–164. https://doi.org/10.1007/s11920-007-0086-7

Dimitropoulos, A., Zyga, O., & Russ, S. W. (2019). Early Social Cognitive Ability in Preschoolers with Prader–Willi Syndrome and Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *49*(11), 4441–4454. https://doi.org/10.1007/s10803-019-04152-4

Downey, D. A., & Knutson, C. L. (1995). Speech and Language Issues. In L. R. Greenswag & R. C. Alexander (Eds.), *Management of Prader-Willi Syndrome* (Second Edi, pp. 142–155). Springer.

Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2014). The inhibitory advantage in bilingual children revisited: Myth or reality? *Experimental Psychology*, *61*(3), 234–251. https://doi.org/10.1027/1618-3169/a000243

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

Dunn, L. M., Dunn, L. M., & Arribas, D. (2006). *Peabody, Test de vocabulario en imágenes*. TEA Ediciones.

Dykens, E. M. (2002). Are jigsaw puzzle skills “spared” in persons with Prader-Willi syndrome? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *43*(3), 343–352. https://doi.org/10.1111/1469-7610.00025

Dykens, E. M., Hodapp, R. M., Walsh, K., & Nash, L. J. (1992). Profiles, Correlates, and Trajectories of Intelligence in Prader-Willi Syndrome. *Journal of the American Academy of Child & Adolescent Psychiatry*, *31*(6), 1125–1130. https://doi.org/https://doi.org/10.1097/00004583-199211000-00022

Dykens, E. M., Lee, E., & Roof, E. (2011). Prader-Willi syndrome and autism spectrum disorders: An evolving story. *Journal of Neurodevelopmental Disorders*, *3*(3), 225–237. https://doi.org/10.1007/s11689-011-9092-5

Dykens, E. M., Roof, E., & Hunt-Hawkins, H. (2017). Cognitive and adaptive advantages of growth hormone treatment in children with Prader-Willi syndrome. *Journal of Child Psychology and Psychiatry*, *58*(1), 64–74. https://doi.org/10.1111/jcpp.12601

Edgin, J. O., Kumar, A., Spanò, G., & Nadel, L. (2011). Neuropsychological effects of second language exposure in Down syndrome. *Journal of Intellectual Disability Research*, *55*(3), 351–356. https://doi.org/10.1111/j.1365-2788.2010.01362.x

Emmorey, K., Luk, G., Pyers, J. E., & Bialystok, E. (2008). The Source of Enhanced Cognitive Control in Bilinguals: Evidence from Bimodal Bilinguals. *Psychological Science*, *19*(12), 1201–1206. https://doi.org/10.1111/j.1467-9280.2008.02224.x

Engel de Abreu, P. M. J., Cruz-Santos, A., & Puglisi, M. L. (2014). Specific language impairment in language-minority children from low-income families. *International Journal of Language and Communication Disorders*, *49*(6), 736–747. https://doi.org/10.1111/1460-6984.12107

Engel de Abreu, P. M. J., Cruz-Santos, A., Tourinho, C. J., Martin, R., & Bialystok, E. (2012). Bilingualism Enriches the Poor: Enhanced Cognitive Control in Low-Income Minority Children. *Psychological Science*, *23*(11), 1364–1371. https://doi.org/10.1177/0956797612443836

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. https://doi.org/10.3758/BF03203267

Esposito, A. G., Baker-Ward, L., & Mueller, S. (2013). Interference Suppression vs. Response Inhibition: An Explanation for the Absence of a Bilingual Advantage in Preschoolers' Stroop Task Performance. *Cognitive Development*, *28*(4), 354–363. https://doi.org/https://doi.org/10.1016/j.cogdev.2013.09.002

Estigarribia, B., Martin, G. E., Roberts, J. E., Spencer, A., Gucwa, A., & Sideris, J. (2011). Narrative skill in boys with fragile X syndrome with and without autism spectrum disorder. *Applied Psycholinguistics*, *32*, 359–388. https://doi.org/10.1017/S0142716410000445

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the Efficiency and Independence of Attentional Networks. *Journal of Cognitive Neuroscience*, *14*(3), 340–347.

Feltmate, K., & Kay-Raining Bird, E. (2008). Language Learning in Four Bilingual Children with Down Syndrome: A Detailed Analysis of Vocabulary and Morphosyntax. *Canadian Journal of Speech-Language Pathology and Audiology*, *32*(1), 6–20.

Fergadiotis, G., Wright, H. H., & Green, S. B. (2015). Psychometric Evaluation of Lexical Diversity Indices: Assessing Lenghth Effects. *Journal of Speech, Language, and Hearing Research*, *58*(3), 840–852. https://doi.org/10.1044/2015_JSLHR-L-14-0280

Field, A. (2009). *Discovering statistics using SPSS (3rd edition).* Sage Publications.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage Publications.

Fiestas, C. E., & Peña, E. D. (2004). Narrative Discourse in Bilingual Children: Language and Task Effects. *Language, Speech, and Hearing Services in Schools*, *35*(2), 155–168. https://doi.org/10.1044/0161-1461(2004/016)

Finestack, L. H., Palmer, M., & Abbeduto, L. (2012). Macrostructural Narrative Language of Adolescents and Young Adults with Down Syndrome or Fragile X Syndrome. *American Journal of Speech-Language Pathology*, *21*(1), 29–46. https://doi.org/10.1044/1058-0360(2011/10-0095)

Foundation for Prader-Willi Research. (n.d.). *About Prader-Willi Syndrome*. Retrieved July 19, 2020, from https://www.fpwr.org/about-prader-willi-syndrome#definition

Friedman, N. P. (2016). Research on individual differences in executive functions: Implications for the bilingual advantage hypothesis. *Linguistic Approaches to Bilingualism*, *6*(5), 535–548. https://doi.org/10.1075/lab.15041.fri

Friesen, D. C., & Bialystok, E. (2012). Metalinguistic Ability in Bilingual Children: The Role of Executive Control. *Rivista Di Psicolinguistica Applicata*, *12*(3), 47–56.

Frost, J. (2017). *Five regression analysis tips to avoid common problems.* https://statisticsbyjim.com/regression/regression-analysis-tips/

Frost, J. (2019). *Overfitting regression models: Problems, detection, and avoidance*. https://statisticsbyjim.com/regression/overfitting-regression-models/

Gabel, S., Tarter, R. E., Gavaler, J., Golden, W. L., Hegedus, A. M., & Maier, B. (1986). Neuropsychological Capacity of Prader-Willi Children: General and Specific Aspects of Impairment. *Applied Research In Mental Retardation*, *7*(4), 459–466. https://doi.org/10.1016/S0270-3092(86)80018-2

Gagarina, N. (2016). Narratives of Russian-German preschool and primary school bilinguals: Rasskaz and Erzaehlung. *Applied Psycholinguistics*, *37*(1), 91–122. https://doi.org/10.1017/S0142716415000430

Gagarina, N., Klop, D., Tsimpli, I. M., & Walters, J. (2016). Narrative abilities in bilingual children. *Applied Psycholinguistics*, *37*(1), 11–17. https://doi.org/10.1017/S0142716415000399

Garayzábal-Heinze, E., Capó, M., Moruno, E., Gonçalves, Ó., Férnandez, M., Lens, M., & Sampaio, A. (2012). Uncommon genetic syndromes and narrative production - Case Studies with Williams, Smith-Magenis and Prader-Willi Syndromes? *International Journal of Developmental Disabilities*, *58*(1), 48–65. https://doi.org/10.1179/2047387711Y.0000000008

García-Alcaraz, E. (2018). The Narrative Abilities of an English-Spanish Bilingual with Prader-Willi Syndrome. *Languages*, *3*(3), 23. https://doi.org/10.3390/languages3030023

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

Gonçalves, Ó., Henriques, M., Alves, A., & Rocha, C. (2001). *Process and Narrative Complexity Evaluation Manual*. Universidade do Minho.

Gonçalves, Ó., Henriques, M., Alves, A., & Soares, L. (2002). Analyzing structure, process and content in narratives of patients diagnosed with agoraphobia. *International Journal of Clinical and Health Psychology*, *2*(3), 389–406.

Gonçalves, Ó., Henriques, M., & Cardoso, G. (2001). *Structure and Narrative Coherence Evaluation Manual*. Universidade do Minho.

Gonçalves, Ó., Henriques, M., Soares, L., & Monteiro, A. (2001). *Narrative Content and Multiplicity Evaluation Manual*. Universidade do Minho.

Gonçalves, Ó., Pinheiro, A. P., Sampaio, A., Sousa, N., Férnandez, M., & Henriques, M. (2011). Autobiographical Narratives in Williams Syndrome: Structural, Process and Content Dimensions. *Journal of Developmental and Physical Disabilities*, *23*, 289–302. https://doi.org/10.1007/s10882-011-9228-2

Gonzalez-Barrero, A. M., & Nadig, A. (2017). Verbal fluency in bilingual children with Autism Spectrum Disorders. *Linguistic Approaches to Bilingualism*, *7*(3/4), 460–475. https://doi.org/10.1075/lab.15023.gon

Gonzalez-Barrero, A. M., & Nadig, A. (2019a). Brief Report: Vocabulary and Grammatical Skills of Bilingual Children with Autism Spectrum Disorders at School Age. *Journal of Autism and Developmental Disorders*, *49*(9), 3888–3897. https://doi.org/10.1007/s10803-019-04073-2

Gonzalez-Barrero, A. M., & Nadig, A. (2019b). Can Bilingualism Mitigate Set-Shifting Difficulties in Children With Autism Spectrum Disorders? *Child Development*, *90*(4), 1043–1060. https://doi.org/10.1111/cdev.12979

Govindarajan, K., & Paradis, J. (2019). Narrative abilities of bilingual children with and without Developmental Language Disorder (SLI): Differentiation and the role of age and input factors. *Journal of Communication Disorders*, *77*, 1–16. https://doi.org/10.1016/j.jcomdis.2018.10.001

Grosjean, F. (1992). Another View of Bilingualism. *Advances in Psychology*, *83*, 51–62. https://doi.org/10.1016/S0166-4115(08)61487-9

Grosjean, F. (2010). *Bilingual: Life and Reality*. Harvard University Press.

Gross-Tsur, V., Landau, Y. E., Benarroch, F., Wertman-Elad, R., & Shalev, R. S. (2001). Cognition, Attention, and Behaviour in Prader-Willy Syndrome. *Journal of Child Neurology*, *16*(4), 288–290. https://doi.org/10.1177/088307380101600411

Grundy, J. G., Chung-Fat-Yim, A., Friesen, D. C., Mak, L., & Bialystok, E. (2017). Sequential congruency effects reveal differences in disengagement of attention for monolingual and bilingual young adults. *Cognition*, *163*, 42–55. https://doi.org/10.1016/j.cognition.2017.02.010

Heilmann, J., Miller, J. F., Nockerts, A., & Dunaway, C. (2010). Properties of the Narrative Scoring Scheme Using Narrative Retells in Young School-Age Children. *American Journal of Speech-Language Pathology*, *19*(2), 154–166. https://doi.org/10.1044/1058-0360(2009/08-0024)

Hermanto, N., Moreno, S., & Bialystok, E. (2012). Linguistic and metalinguistic outcomes of intense immersion education: how bilingual? *International Journal of Bilingual Education and Bilingualism*, *15*(2), 131–145. https://doi.org/10.1080/13670050.2011.652591

Hernández, M., Costa, A., Fuentes, L. J., Vivas, A. B., & Sebastián-Gallés, N. (2010). The impact of bilingualism on the executive control and orienting networks of attention. *Bilingualism: Language and Cognition*, *13*(3), 315–325. https://doi.org/10.1017/S1366728909990010

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin and Review*, *18*(4), 625–658. https://doi.org/10.3758/s13423-011-0116-7

Hipfner-Boucher, K., Milburn, T., Weitzman, E., Greenberg, J., Pelletier, J., & Girolametto, L. (2015). Narrative abilities in subgroups of English language learners and monolingual peers. *International Journal of Bilingualism*, *19*(6), 677–692. https://doi.org/10.1177/1367006914534330

Hoang, H., Gonzalez-Barrero, A. M., & Nadig, A. (2018). Narrative Skills of Bilingual Children with Autism Spectrum Disorder. *Discours. Revue de Linguistique, Psycholinguistique et Informatique*, *23*. https://doi.org/10.4000/discours.9856

Hodapp, A. F., & Gerken, K. C. (1999). Correlations between scores for Peabody Picture Vocabulary Test-III and the Wechsler Intelligence Scale for Children-III. *Psychological Reports*, *84*, 1139–1142. https://doi.org/10.2466/pr0.84.3.1139-1142

IBM Corp. (2018). *IBM SPSS Statistics (version 25) [Computer Software]*.

Iluz-Cohen, P., & Walters, J. (2012). Telling stories in two languages: Narratives of bilingual preschool children with typical and impaired language. *Bilingualism: Language and Cognition*, *15*(1), 58–74. https://doi.org/10.1017/S1366728911000538

INEE. (2010). *Evaluación General de Diagnóstico 2009. Educación Primaria. Cuarto Curso. Informe De Resultados.* Ministerio de eduación. https://www.educacionyfp.gob.es/inee/dam/jcr:c3d1762a-0d7b-4b26-b56e-a10a12870938/pdf-completo-informe-egd-2009.pdf

INEE. (2011). *Evaluación General de Diagnóstico 2010: Educación Secundaria Obligatoria. Segundo Curso. Informe de resultados.* Ministerio de eduación. https://www.educacionyfp.gob.es/inee/dam/jcr:134d0cfd-1090-4c21-8b10-66a6b07b565a/informe-egd-2010.pdf

Ivanova, I., Salmon, D. P., & Gollan, T. H. (2014). Which Language Declines More? Longitudinal versus Cross-sectional Decline of Picture Naming in Bilinguals with Alzheimer's Disease. *Journal of the International Neuropsychological Society*, *20*(5), 534–546. https://doi.org/10.1017/S1355617714000228

Jarmendi, A. (2002). *Fragile X and Bilingualism. A case study* [Master's thesis, The University of Szeged]. http://diploma.bibl.u-szeged.hu/76332/

JASP Team. (2019). *JASP (Version 0.11.1) [Computer software]*.

Jauregui, J., Arias, C., Vegas, O., Alén, F., Martinez, S., Copet, P., & Thuilleaux, D. (2007). A neuropsychological assessment of frontal cognitive functions in Prader-Willi syndrome. *Journal of Intellectual Disability Research*, *51*(5), 350–365. https://doi.org/10.1111/j.1365-2788.2006.00883.x

Jevtovi, M., Duñabeitia, J. A., & De Bruin, A. (2019). How do bilinguals switch between languages in different interactional contexts? A comparison between voluntary and mandatory language switching. *Bilingualism: Language and Cognition*, *23*, 401–413. https://doi.org/10.1017/S1366728919000191

Johnson, S. G. (n.d.). *The Nlopt nonlinear-optimization package*. http://github.com/stevengj/nlopt

Johnson, W. (1944). Studies in language behavior: A program of research. *Psychological Monographs*, *56*(2), 1–15.

Kalanthroff, E., & Henik, A. (2013). Individual but not fragile: Individual differences in task control predict Stroop facilitation. *Consciousness and Cognition*, *22*(2), 413–419. https://doi.org/10.1016/j.concog.2013.01.010

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin and Review*, *9*(4), 637–671. https://doi.org/10.3758/BF03196323

Kapa, L. L., & Colombo, J. (2013). Attentional control in early and later bilingual children. *Cognitive Development*, *28*(3), 233–246. https://doi.org/10.1016/j.cogdev.2013.01.011

Kapalková, S., Polišenská, K., Marková, L., & Fenton, J. (2016). Narrative abilities in early successive bilingual Slovak-English children: A cross-language comparison. *Applied Psycholinguistics*, *37*(1), 145–164. https://doi.org/10.1017/S0142716415000454

Karam, S. M., Riegel, M., Segal, S. L., Félix, T. M., Barros, A. J. D., Santos, I. S., Matijasevich, A., Giugliani, R., & Black, M. (2015). Genetic Causes of Intellectual Disability in a Birth Cohort: A Population-Based Study. *American Journal of Medical Genetics, Part A*, *167*(6), 1204–1214. https://doi.org/10.1002/ajmg.a.37011

Kartushina, N., & Martin, C. D. (2019). Third-language learning affects bilinguals' production in both their native languages: A longitudinal study of dynamic changes in L1, L2 and L3 vowel production. *Journal of Phonetics*, *77*, 100920. https://doi.org/10.1016/j.wocn.2019.100920

Katsarou, D., & Andreou, G. (2019). Bilingualism in Down Syndrome: A Greek Study. *International Journal of Disability, Development and Education*, 1–7. https://doi.org/10.1080/1034912X.2019.1684458

Kay-Raining Bird, E., Genesee, F., & Verhoeven, L. (2016). Bilingualism in children with developmental disorders: A narrative review. *Journal of Communication Disorders*, *63*, 1–14. https://doi.org/10.1016/j.jcomdis.2016.07.003

Kay-Raining Bird, E., Lamond, E., & Holden, J. (2012). Survey of bilingualism in autism spectrum disorders. *International Journal of Language and Communication Disorders*, *47*(1), 52–64. https://doi.org/10.1111/j.1460-6984.2011.00071.x

Kay-Raining Bird, E., Trudeau, N., Thordardottir, E., Sutton, A., & Thorpe, A. (2005). The Language Abilities of Bilingual Children with Down Syndrome. *American Journal of Speech-Language Pathology*, *14*(3), 187–199. https://doi.org/10.1044/1058-0360(2005/019)

Kizach, J. (2014). *Analyzing Likert-scale data with mixed-effects linear models: a simulation study*. [Poster presentation]. Linguistic Evidence 2014, University of Tübingen, Germany.

Kleppe, S. A., Katayama, K. M., Shipley, K., & Foushee, D. (1990). The Speech and Language Characteristics of Children with Prader-Willi Syndrome. *Journal of Speech and Hearing Disorders*, *55*, 300–309. https://doi.org/https://doi.org/10.1044/jshd.5502.300

Kolinsky, R., Cary, L., & Morais, J. (1987). Awareness of words as phonological entities: The role of literacy. *Applied Psycholinguistics*, *8*(3), 223–232. https://doi.org/10.1017/S0142716400000278

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kousaie, S., & Phillips, N. A. (2012a). Ageing and bilingualism: Absence of a “bilingual advantage” in Stroop interference in a nonimmigrant sample. *The Quarterly Journal of Experimental Psychology*, *65*(2), 356–369. https://doi.org/10.1080/17470218.2011.604788

Kousaie, S., & Phillips, N. A. (2012b). Conflict monitoring and resolution: Are two languages better than one? Evidence from reaction time and event-related brain potentials. *Brain Research*, *1446*, 71–90. https://doi.org/10.1016/j.brainres.2012.01.052

Krasileva, K. E., Sanders, S. J., & Bal, V. H. (2017). Peabody Picture Vocabulary Test: Proxy for Verbal IQ in Genetic Studies of Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *47*(4), 1073–1085. https://doi.org/10.1007/s10803-017-3030-7

Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*, *25*(5), 497–514. https://doi.org/10.1080/20445911.2013.799170

Kunnari, S., Välimaa, T., & Laukkanen-Nevala, P. (2016). Macrostructure in the narratives of monolingual Finnish and bilingual Finnish-Swedish children. *Applied Psycholinguistics*, *37*(1), 123–144. https://doi.org/10.1017/S0142716415000442

Kurvers, J., van Hout, R., & Vallen, T. (2006). Discovering Features of Language: Metalinguistic Awareness of Adult Illiterates. In I. Van de Craats, J. Kurvers, & M. Young-Scholten (Eds.), *Low-Educated Adult Second Language and Literacy Learners* (pp. 69–88). LOT Occasional Series 6.

Laloi, A., de Jong, J., & Baker, A. (2017). Can executive functioning contribute to the diagnosis of SLI in bilingual children?: A study on response inhibition. *Linguistic Approaches to Bilingualism*, *7*(3–4), 431–459. https://doi.org/10.1075/lab.15020.lal

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The Sources of Four Commonly Reported Cutoff Criteria. What Did They Really Say? *Organizational Research Methods*, *9*(2), 202–220. https://doi.org/10.1177/1094428105284919

Leppink, J., Winston, K., & O'Sullivan, P. (2016). Statistical significance does not imply a real effect. *Perspectives on Medical Education*, *5*(2), 122–124. https://doi.org/10.1007/s40037-016-0256-6

Lewis, B. A. (2006). Speech and Language Disorders Associated with Prader-Willi Syndrome. In M. G. Butler, P. D. K. Lee, & B. Y. Whitman (Eds.), *Management of Prader-Willi Syndrome* (Third Edit, pp. 272–283). Springer-Verlag. https://doi.org/10.1007/978-0-387-33536-0_9

Lewis, B. A., Freebairn, L., Heeger, S., & Cassidy, S. B. (2002). Speech and Language Skills of Individuals with Prader-Willi Syndrome. *American Journal of Speech-Language Pathology*, *11*(3), 285–294. https://doi.org/10.1044/1058-0360(2002/033)

Liceras, J. M., Fernández-Fuertes, R., & Klassen, R. (2016). Language dominance and language nativeness: The view from English-Spanish code-switching. In R. E. Guzzardo Tamargo, C. M. Mazak, & M. C. Parafita Couto (Eds.), *Spanish-English codeswitching in the Caribbean and the US* (pp. 107–138). John Benjamins. https://doi.org/10.1075/ihll.11.05lic

Liceras, J. M., & García-Alcaraz, E. (2019). Grammatical Gender in Atypical Language Development : A Case Study on the Interpretation and Production of Sentence Internal Codeswitching by an English – Spanish bilingual with Prader – Willi Syndrome. *Journal of Monolingual and Bilingual Speech*, *1*(2), 225–247. https://doi.org/10.1558/jmbs.v1i2.11878

Linck, J. A., & Cunnings, I. (2015). The Utility and Application of Mixed-Effects Models in Second Language Research. *Language Learning*, *65*(S1), 185–207. https://doi.org/10.1111/lang.12117

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, *91*(3), 295–327. https://doi.org/10.1037/0033-295X.91.3.295

Luk, G., Anderson, J. A. E., Craik, F. I. M., Grady, C., & Bialystok, E. (2010). Distinct neural correlates for two types of inhibition in bilinguals: Response inhibition versus interference suppression. *Brain and Cognition*, *74*(3), 347–357. https://doi.org/10.1016/j.bandc.2010.09.004

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, *25*(5), 605–621. https://doi.org/10.1080/20445911.2013.795574

Luk, G., De Sa, E., & Bialystok, E. (2011). Is there a relation between onset age of bilingualism and enhancement of cognitive control? *Bilingualism: Language and Cognition*, *14*(4), 588–595. https://doi.org/10.1017/S1366728911000010

MacLeod, C. M. (2006). Stroop Effect in Language. *Encyclopedia of Language & Linguistics*, 161–165. https://doi.org/10.1016/b0-08-044854-2/00871-3

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk. 3rd Edition.* Lawrence Erlbaum Associates. https://doi.org/10.21415/3mhn-0z89

Mainly, T., Robertson, I. H., Anderson, V., & Nimmo-Smith, I. (1999). *The test of everyday attention for children (TEA-CH)*. Thames Valley Test Company.

Marian, V., Blumenfeld, H. K., Mizrahi, E., Kania, U., & Cordes, A. K. (2013). Multilingual Stroop performance: Effects of trilingualism and proficiency on inhibitory control. *International Journal of Multilingualism*, *10*(1), 82–104. https://doi.org/10.1080/14790718.2012.708037

Marian, V., & Shook, A. (2012). The cognitive benefits of being bilingual. *Cerebrum: The Dana Forum on Brain Science*, *13*.

Marinova-Todd, S. H., Colozzo, P., Mirenda, P., Stahl, H., Kay-Raining Bird, E., Parkington, K., Cain, K., Scherba de Valenzuela, J., Segers, E., MacLeod, A. A. A. N., & Genesee, F. (2016). Professional practices and opinions about services available to bilingual children with developmental disabilities: An international study. *Journal of Communication Disorders*, *63*, 47–62. https://doi.org/10.1016/j.jcomdis.2016.05.004

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

Martin, S. (2017). *Bilingual Outcomes for a Child with Down Syndrome in the Early French Immersion Program* [Master's thesis, Dalhousie University]. https://dalspace.library.dal.ca/handle/10222/73509

Marton, K., Gehebe, T., & Pazuelo, L. (2019). Cognitive Control along the Language Spectrum: From the Typical Bilingual Child to Language Impairment. *Seminars in Speech and Language*, *40*(4), 256–271. https://doi.org/10.1055/s-0039-1692962

Mayer, M. (1967). *A boy, a dog and a frog*. Dial Press.

Mayer, M. (1969). *Frog, where are You?* Dial Press.

McFadden, T. U., & Gillam, R. B. (1996). An Examination of the Quality of Narratives Produced by Children with Language Disorders. *Language, Speech, and Hearing Services in Schools*, *27*(1), 48–56. https://doi.org/10.1044/0161-1461.2701.48

Meir, N., & Novogrodsky, R. (2019). Prerequisites of Third-Person Pronoun Use in Monolingual and Bilingual Children With Autism and Typical Language Development. *Frontiers in Psychology*, *10*, 2289. https://doi.org/10.3389/fpsyg.2019.02289

Mellor, A. (2011). Essay Length, Lexical Diversity and Automatic Essay Scoring. *Memoirs of the Osaka Institute of Technology*, *55*(2), 1–14.

Montgomery, D. E., & Koeltzow, T. E. (2010). A review of the day-night task: The Stroop paradigm and interference control in young children. *Developmental Review*, *30*(3), 308–330. https://doi.org/10.1016/j.dr.2010.07.001

Montrul, S. (2008). *Incomplete acquisition in bilingualism: re-examining the age factor*. John Benjamins. https://doi.org/10.1075/sibil.39

Moreno, S., Bialystok, E., Wodniecka, Z., & Alain, C. (2010). Conflict resolution in sentence processing by bilinguals. *Journal of Neurolinguistics*, *23*(6), 564–579. https://doi.org/10.1016/j.jneuroling.2010.05.002

Myers, M. (1981). *A Procedure for Writing Assessment and Holistic Scoring*. National Council of Teachers of English.

Nadig, A., & Gonzalez-Barrero, A. M. (2019). Proficient bilingualism may alleviate some executive function difficulties in children with Autism Spectrum Disorders. In I. A. Sekerina, L. Spradlin, & V. Valian (Eds.), *Bilingualism, Executive Function, and Beyond: Questions and insights* (pp. 337–354). John Benjamins. https://doi.org/https://doi.org/10.1075/sibil.57.20nad

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*, 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*, 353–383. https://doi.org/10.1016/0010-0285(77)90012-3

Nelson, R. L., Damico, J. S., & Smith, S. K. (2008). Applying eye movement miscue analysis to the reading patterns of children with language impairment. *Clinical Linguistics and Phonetics*, *22*(4–5), 293–303. https://doi.org/10.1080/02699200801919265

Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education*, *15*(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y

Nugnes, R., Zito, E., Mozzillo, E., Camarca, M. E., Riccio, M. P., Terrone, G., Melis, D., Bravaccio, C., & Franzese, A. (2013). Good cognitive performances in a child with Prader-Willi syndrome. *Italian Journal of Pediatrics*, *39*, 74. https://doi.org/10.1186/1824-7288-39-74

Otheguy, R., & Lapidus, N. (2005). Matización de la teoría de la simplificación en las lenguas en contacto: El concepto de la adaptación en el español de Nueva York. In L. Ortiz & M. Lacorte (Eds.), *Contactos y contextos lingüísticos: El español en Estados Unidos y en contacto con otras lenguas* (pp. 143–160). Editorial Iberoamericana/Vervuert. https://doi.org/10.31819/9783865278586-009

Paap, K. R. (2014). The role of componential analysis, categorical hypothesising, replicability and confirmation bias in testing for bilingual advantages in executive functioning. *Journal of Cognitive Psychology*, *26*(3), 242–255. https://doi.org/10.1080/20445911.2014.891597

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*(2), 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., & Liu, Y. (2014). Conflict resolution in sentence processing is the same for bilinguals and monolinguals: The role of confirmation bias in testing for bilingual advantages. *Journal of Neurolinguistics*, *27*(1), 50–74. https://doi.org/10.1016/j.jneuroling.2013.09.002

Papagno, C., & Vallar, G. (1995). To Learn or Not to Learn : Vocabulary in Foreign Languages and the Problem with Phonological Memory. *Broken Memories: Case Estudies in Memory Impairment*, 334–343.

Paradis, J. (2016). An agenda for knowledge-oriented research on bilingualism in children with developmental disorders. *Journal of Communication Disorders*, *63*, 79–84. https://doi.org/10.1016/j.jcomdis.2016.08.002

Paradis, J., & Govindarajan, K. (2018). Bilingualism and children with developmental language and communication disorders. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual Cognition and Language: The state of the science across its subfields* (pp. 347–370). John Benjamins. https://doi.org/10.1075/sibil.54.16par

Parker, M. D., & Brorson, K. (2005). A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). *First Language*, *25*(3), 365–376. https://doi.org/10.1177/0142723705059114

Peal, E., & Lambert, W. E. (1962). The Relation of Bilingualism to Intelligence. *Psychological Monographs: General and Applied*, *76*(27), 1–23. https://doi.org/10.1037/h0093840

Pearson, B. Z. (2002). Narrative Competence among Monolingual and Bilingual School Children in Miami. In K. Oller & R. Eilers (Eds.), *Language and Literacy in Bilingual Children* (pp. 135–174). Multilingual Matters. https://doi.org/10.21832/9781853595721-008

Pegoraro, L. F. L., Steiner, C. E., Celeri, E. H. R. V., Banzato, C. E. M., & Dalgalarrondo, P. (2014). Cognitive and behavioral heterogeneity in genetic syndromes. *Jornal de Pediatria*, *90*(2), 155–160. https://doi.org/10.1016/j.jped.2013.06.010

Pelham, S. D., & Abrams, L. (2014). Cognitive Advantages and Disadvantages in Early and Late bilinguals. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*(2), 313–325. https://doi.org/10.1037/a0035224

Peña, E. D. (2016). Supporting the home language of bilingual children with developmental disabilities: From knowing to doing. *Journal of Communication Disorders*, *63*, 85–92. https://doi.org/10.1016/j.jcomdis.2016.08.001

Peristeri, E., Baldimtsi, E., Andreou, M., & Tsimpli, I. M. (2020). The impact of bilingualism on the narrative ability and the executive functions of children with autism spectrum disorders. *Journal of Communication Disorders*, *85*, 105999. https://doi.org/10.1016/j.jcomdis.2020.105999

Peristeri, E., Baldimtsi, E., Durrleman, S., & Tsimpli, I. M. (2019). Bilingualism Effects in Children with Developmental Language Disorder: Metalinguistic Awareness, Executive Functions, and False-Belief Reasoning. In M. M. Brown & B. Dailey (Eds.), *Proceedings of the 43rd Boston University Conference on Language Development* (pp. 523–536). Cascadilla Press.

Perovic, A., & Lochet, S. (2015). *Grammatical morphology in bilingual Williams syndrome: A single case study*. [Paper presentation]. Bi-SLI 2015, Tours, France.

Pesco, D., MacLeod, A. A. N., Kay-Raining Bird, E., Cleave, P. L., Trudeau, N., Scherba de Valenzuela, J., Cain, K., Marinova-Todd, S. H., Colozzo, P., Stahl, H., Segers, E., & Verhoeven, L. (2016). A multi-site review of policies affecting opportunities for children with developmental disabilities to become bilingual. *Journal of Communication Disorders*, *63*, 15–31. https://doi.org/10.1016/j.jcomdis.2016.05.008

Phillips, N. D. (2016). *YaRrr! The Pirate's Guide to R*. http://www.thepiratesguidetor.com.

Pinto-Cardona, E. I. (2017). *Effects of Bilingualism in Short-Term Memory in Individuals with Down Syndrome* [Master's thesis, Florida International University]. https://doi.org/10.25148/etd.FIDC001971

Plataforma per la llengua, P. (2019). *Estudi sociolingüístic als patis d'escoles i instituts de zones urbanes de Catalunya*. https://www.plataforma-llengua.cat/que-fem/estudis-i-publicacions/248/estudi-sociolinguistic-als-patis-descoles-i-instituts-de-zones-urbanes-de-catalunya

Poarch, G. J., & Bialystok, E. (2015). Bilingualism as a model for multitasking. *Developmental Review*, *35*, 113–124. https://doi.org/10.1016/j.dr.2014.12.003

Polo, S. L. (2017). *Teaching English as a Foreign Language to Students with Down Syndrome : A case Study* [Master's thesis, Universidad Complutense de Madrid]. https://eprints.ucm.es/id/eprint/49726/

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Poulin-Dubois, D., Blaye, A., Coutya, J., & Bialystok, E. (2011). The effects of bilingualism on toddlers' executive functioning. *Journal of Experimental Child Psychology*, *108*(3), 567–579. https://doi.org/10.1016/j.jecp.2010.10.009

R Core Team. (2019). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Viena, Austria. https://www.r-project.org/.

Rezzonico, S., Chen, X., Cleave, P. L., Greenberg, J., Hipfner-Boucher, K., Johnson, C. J., Milburn, T., Pelletier, J., Weitzman, E., & Girolametto, L. (2015). Oral narratives in monolingual and bilingual preschoolers with SLI. *International Journal of Language and Communication Disorders*, *50*(6), 830–841. https://doi.org/10.1111/1460-6984.12179

Rezzonico, S., Goldberg, A., Ka-Yan Mak, K., Yap, S., Milburn, T., Belletti, A., & Girolametto, L. (2016). Narratives in Two Languages: Storytelling of Bilingual Cantonese-English Preschoolers. *Journal of Speech, Language, and Hearing Research*, *59*, 521–532. https://doi.org/10.1044/2015_JSLHR-L-15-0052

Ricciardelli, L. A. (1992). Bilingualism and Cognitive Development in Relation to Threshold Theory. *Journal of Psycholinguistic Research*, *21*(4), 301–316. https://doi.org/10.1007/BF01067515

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, *6*(2), 135–147. https://doi.org/10.1016/j.edurev.2010.12.001

Roberts, J. E., Price, J., & Malkin, C. (2007). Language and Communication Development in Down Syndrome. *Mental Retardation and Developmental Disabilities, Research Reviews*, *13*, 26–35. https://doi.org/10.1002/mrdd.20136

Roehr-Brackin, K. (2018). *Metalinguistic awareness and second language acquisition*. https://doi.org/10.4324/9781315661001

Rojas, R., & Iglesias, A. (2009). Making a Case for Language Sampling. *The ASHA Leader*, *14*(3), 10–13. https://doi.org/10.1044/leader.FTR1.14032009.10

Roof, E., Stone, W., MacLean, W., Feurer, I. D., Thompson, T., & Butler, M. G. (2000). Intellectual characteristics of Prader-Willi syndrome: Comparison of genetic subtypes. *Journal of Intellectual Disability Research*, *44*(1), 25–30. https://doi.org/10.1046/j.1365-2788.2000.00250.x

RStudio Team. (2019). *RStudio: Integrated Development Environment for R.* RStudio, PBC, Boston, MA. https://rstudio.com/

Ryskin, R. A., Brown-Schmidt, S., Canseco-Gonzalez, E., Yiu, L. K., & Nguyen, E. T. (2014). Visuospatial perspective-taking in conversation and the role of bilingual experience. *Journal of Memory and Language*, *74*, 46–76. https://doi.org/10.1016/j.jml.2014.04.003

SALT Software LCC. (2016). *NSS Rubric for Frog Where are You? by Mercer Mayer (1969) [PDF file]*. https://www.saltsoftware.com/media/wysiwyg/codeaids/NSS_FWAY_Rubric.pdf

SALT Software LCC. (2017). *NSS Scoring Guide [PDF file]*. https://www.saltsoftware.com/media/wysiwyg/codeaids/NSS_Scoring_Guide.pdf

Schneider, P., Dubé, R. V., & Hayward, D. (2005). *The Edmonton Narrative Norms Instrument (ENNI)*. http://www.rehabresearch.ualberta.ca/enni/.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002a). *E-Prime: User's guide*. Psychology Software Incorporated.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002b). *E-Prime reference guide*. Psychology Software Incorporated.

Sekerina, I. A., Spradlin, L., & Valian, V. (2019). Bilingualism, executive function, and beyond: Questions and insights. In I. A. Sekerina, L. Spradlin, & V. Valian (Eds.), *Bilingualism, Executive Function, and Beyond: Questions and Insights*. John Benjamins. https://doi.org/https://doi.org/10.1075/sibil.57.01sek

Simon, J. R., & Rudell, A. (1967). Auditory SR compatibility: The effect of an irrelevant directional cue on information processing. *Journal of Applied Psychology*, *51*(3), 300–304.

Slobin, D. I. (2005). Relating Narrative Events in Translation. In D. D. Ravid & H. B.-Z. Shyldkrot (Eds.), *Perspectives on Language and Language Development* (pp. 115–129). Springer. https://doi.org/10.1007/1-4020-7911-7_10

Squires, K. E., Lugo-Neris, M. J., Peña, E. D., Bedore, L. M., Bohman, T. M., & Gillam, R. B. (2014). Story retelling by bilingual children with language impairments and typically developing controls. *International Journal of Language and Communication Disorders*, *49*(1), 60–74. https://doi.org/10.1111/1460-6984.12044

Stauder, J. E. A., Boer, H., Gerits, R. H. A., Tummers, A., Whittington, J., & Curfs, L. M. (2005). Differences in behavioural phenotype between parental deletion and maternal uniparental disomy in Prader-Willi syndrome: An ERP study. *Clinical Neurophysiology*, *116*(6), 1464–1470. https://doi.org/10.1016/j.clinph.2005.02.019

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary (3rd ed.)*. Oxford University Press.

Strömqvist, S., & Verhoeven, L. (2004). *Relating events in narrative Volume 2: Typological and Contextual Perspectives*. Psychology Press. https://doi.org/10.4324/9781410609694

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Teubner-Rhodes, S. E., Mishler, A., Corbett, R., Andreu, L., Sanz-Torrent, M., Trueswell, J. C., & Novick, J. M. (2016). The effects of bilingualism on conflict monitoring, cognitive control, and garden-path recovery. *Cognition*, *150*, 213–231. https://doi.org/10.1016/j.cognition.2016.02.011

Traverso, L., Fontana, M., Usai, M. C., & Passolunghi, M. C. (2018). Response Inhibition and Interference Suppression in Individuals with Down Syndrome Compared to Typically Developing Children. *Frontiers in Psychology*, *9*, 660. https://doi.org/10.3389/fpsyg.2018.00660

Treffers-Daller, J. (2019). What Defines Language Dominance in Bilinguals? *Annual Review of Linguistics*, *5*, 375–393. https://doi.org/10.1146/annurev-linguistics-011817-045554

Trudeau, N., Kay-Raining Bird, E., Sutton, A., & Cleave, P. L. (2011). Développement lexical chez les enfants bilingues avec Trisomie 21. *Enfance*, *2011*(3), 383–404. https://doi.org/10.4074/S0013754511003089

Tsimpli, I. M., Peristeri, E., & Andreou, M. (2016). Narrative production in monolingual and bilingual children with specific language impairment. *Applied Psycholinguistics*, *37*(1), 195–216. https://doi.org/10.1017/S0142716415000478

Turell, M. T., & Moyer, M. G. (2008). Transcription. In L. Wei & M. G. Moyer (Eds.), *The Blackwell Guide to Research Methods in Bilingualism and Multilingualism*. Blackwell Publishing Ltd. https://doi.org/10.1002/9781444301120.ch11

Valian, V. (2015). Bilingualism and cognition. *Bilingualism: Language and Cognition*, *18*(1), 3–24. https://doi.org/10.1017/S1366728914000522

Vallar, G., & Papagno, C. (1993). Preserved Vocabulary Acquisition in Down's Syndrome: The Role of Phonological Short-term Memory. *Cortex*, *29*(3), 467–483. https://doi.org/10.1016/S0010-9452(13)80254-7

Van Borsel, J., Defloor, T., & Curfs, L. M. (2007). Expressive Language in Persons with Prader-Willi Syndrome. *Genetic Counseling*, *18*(1), 17–28.

Van der Linden, L., Verreyt, N., De Letter, M., Hemelsoet, D., Mariën, P., Santens, P., Stevens, M., Szmalec, A., & Duyck, W. (2018). Cognate effects and cognitive control in patients with parallel and differential bilingual aphasia. *International Journal of Language and Communication Disorders*, *53*(3), 515–525. https://doi.org/10.1111/1460-6984.12365

Veltman, M. W. M., Thompson, R. J., Roberts, S. E., Thomas, N. S., Whittington, J., & Bolton, P. F. (2004). Prader-Willi syndrome: A study comparing deletion and uniparental disomy cases with reference to autism spectrum disorders. *European Child and Adolescent Psychiatry*, *13*(1), 42–50. https://doi.org/10.1007/s00787-004-0354-6

Verhoeven, L., & Strömqvist, S. (2001). *Narrative development in a multilingual context*. John Benjamins. https://doi.org/https://doi.org/10.1075/sibil.23

Vinerte, S. (2018). *Effects of Bilingualism on Cognitive Control: Considering the Age of Immersion and Different Linguistic Environments* [Doctoral Dissertation, University of Ottawa]. https://ruor.uottawa.ca/handle/10393/38063

Vissers, L. E. L. M., Gilissen, C., & Veltman, J. A. (2016). Genetic studies in intellectual disability and related disorders. *Nature Reviews Genetics*, *17*(1), 9–18. https://doi.org/10.1038/nrg3999

Walley, R. M., & Donaldson, M. D. C. (2005). An investigation of executive function abilities in adults with Prader-Willi syndrome. *Journal of Intellectual Disability Research*, *49*(8), 613–625. https://doi.org/10.1111/j.1365-2788.2005.00717.x

Wang, L., Li, J., Shuang, M., Lu, T., Wang, Z., Zhang, T., Yue, W., Jia, M., Ruan, Y., Liu, J., Wu, Z., Zhang, D., & Wang, L. (2018). Association study and mutation sequencing of genes on chromosome 15q11-q13 identified GABRG3 as a susceptibility gene for autism in Chinese Han population. *Translational Psychiatry*, *8*, 152. https://doi.org/10.1038/s41398-018-0197-4

Wechsler, D. (1991). *Manual for the Wechsler Intelligence Scale for Children-Third Edition (WISC-III)*. The Psychological Corporation.

Wey, E., Bartholdi, D., Riegel, M., Nazlican, H., Horsthemke, B., Schinzel, A., & Baumer, A. (2005). Mosaic imprinting defect in a patient with an almost typical expression of the Prader-Willi syndrome. *European Journal of Human Genetics*, *13*(3), 273–277. https://doi.org/10.1038/sj.ejhg.5201337

Whitman, B. Y., & Greenswag, L. R. (1995). Psychological and Behavioral Management. In L. R. Greenswag & R. C. Alexander (Eds.), *Management of Prader-Willi Syndrome* (Second Edi, pp. 125–141). Springer-Verlag.

Whitman, B. Y., & Thompson, T. (2006). Neurodevelopmental and Neuropsychological Aspects of Prader-Willi Syndrome. In M. Butler, P. D. K. Lee, & B. Y. Whitman (Eds.), *Management of Prader-Willi Syndrome: Third Edition* (Third Edit, pp. 245–271). Springer New York. https://doi.org/https://doi.org/10.1007/978-0-387-33536-0_8

Whittington, J., & Holland, A. (2017). Cognition in people with Prader-Willi syndrome: Insights into genetic influences on cognitive and social development. *Neuroscience and Biobehavioral Reviews*, *72*, 153–167. https://doi.org/10.1016/j.neubiorev.2016.09.013

Wiig, E. H., Semel, E., & Secord, W. A. (2018). *CELF-5, Evaluación Clínica de los Fundamentos del Lenguaje*. Pearson Assessments.

Wilson, B. A., Alderman, N., Burgess, P. W., Emslie, H. E., & Evans, J. J. (1996). *Behavioral Assessment of the Dysexecutive Syndrome*. Thames Valley Test Company.

Winter, B. (2011). Pseudoreplication in Phonetic Research. *ICPhS XVII*, 2137–2140.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. In *arXiv: 1308.5499*. https://arxiv.org/pdf/1308.5499.pdf

Winter, B. (2019). *Statistics for linguists: An introduction using R.* Routledge.

Woll, B., & Grove, N. (1996). On Language Deficits and Modality in Children With Down Syndrome: A Case Study of Twins Bilingual in BSL and English. *Journal of Deaf Studies and Deaf Education*, *1*(4), 271–278. https://doi.org/10.1093/oxfordjournals.deafed.a014302

Woodcock, K. A., Oliver, C., & Humphreys, G. (2009). Task-switching deficits and repetitive behaviour in genetic neurodevelopmental disorders: Data from children with Prader - Willi syndrome chromosome 15 q11 - q13 deletion and boys with Fragile X syndrome. *Cognitive Neuropsychology*, *26*(2), 172–194. https://doi.org/10.1080/02643290802685921

Yang, S., Yang, H., & Lust, B. (2011). Early childhood bilingualism leads to advances in executive attention: Dissociating culture and language. *Bilingualism: Language and Cognition*, *14*(3), 412–422. https://doi.org/10.1017/S1366728910000611

Yelland, G. W., Pollard, J., & Mercuri, A. (1993). The metalinguistic benefits of limited contact with a second language. *Applied Psycholinguistics*, *14*(4), 423–444. https://doi.org/10.1017/S0142716400010687

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): a method of assessing executive function in children. *Nature Protocols, 1*(1), 297–301. https://doi.org/10.1038/nprot.2006.46

Zevenbergen, A. A., Whitehurst, G. J., & Zevenbergen, J. A. (2003). Effects of a shared-reading intervention on the inclusion of evaluative devices in narratives of children from low-income families. *Journal of Applied Developmental Psychology*, *24*(1), 1–15. https://doi.org/10.1016/S0193-3973(03)00021-2

Appendix A. Participants' individual characteristics I and II

Participant	Gender	Group	PWS type	Maternal studies	Chronological age	Speaking (AoO)[a] SP[b]	Reading (AoO) SP	Writing (AoO) SP	Speaking (AoO) CA[c]	Reading (AoO) CA	Writing (AoO) CA
M101	female	monoling.	deletion	university	47;0	2.5	5	6			
M102	male	monoling.	deletion	university	22;5	3	6	6			
M103	female	monoling.	imprinting	university	31;6	2.5	7.5	7.5			
M104	female	monoling.	deletion	secondary school	15;9	1.5	8	8			
M105	male	monoling.	deletion	university	19;10	2	4	4			
M106	male	monoling.	deletion	university	9;4	4	5	5			
M107	male	monoling.	deletion	secondary school	20;0	5	6	6			
M108	male	monoling.	deletion	postsecondary	19;5	2	8	8			
B201	female	bilingual	deletion	university	10;6	3	7.5	6	3	7.5	6
B202	male	bilingual	deletion	postsecondary	10;5	4	6	6	4	6	6
B203	male	bilingual	disomy	university	15;2	6	6	6	2	5	5
B204	female	bilingual	deletion	university	15;2	6	7	7	3	5	5
B205	female	bilingual	imprinting	university	28;0	1.5	5	5	1.5	5	5
B206	male	bilingual	disomy	university	15;3	4	12	12	3	12	12
B207	male	bilingual	deletion	university	33;10	4.5	6	6	4.5	6	6

[a] Age of onset (AoO).
[b] Spanish (SP).
[c] Catalan (CA).

Participant	Oral Com.[a] SP	Oral Pro. SP	Written Com. SP	Written Pro. SP	Oral Com. CA	Oral Pro. CA	Written Com. CA	Written Pro. CA	Nonverbal IQ[b]	Recep. Vocab. SP	Recep. Vocab. SP Age[c]	Sent. Recall SP	Sent. Recall SP Age	Recep. Vocab. CA	Recep. Vocab. CA Age	Sent. Recall CA	Sent. Recall CA Age
M101	2.88	2.86	1.50	5.33					86	55	11;2	55	4;5				
M102	3.50	2.29	4.13	4.00					55	55	8;3	60	5;3				
M103	4.13	6.00	5.13	7.33					87	55	12;6	100	15;1				
M104	4.50	2.69	5.38	6.00					69	55	9;8	65	6;6				
M105	3.75	2.43	2.75	4.00					74	94	17;2	75	7;8				
M106	2.50	2.57	3.00	6.33					79	89	7;11	65	4;1				
M107	3.88	5.29	4.57	5.33					81	55	10;7	60	5;3				
M108	3.14	4.18	6.43	6.83					75	55	7;5	55	3;0				
B201	3.25	3.92	5.75	6.67	1.38	1.85	3.50	5.33	112	63	7;5	55	3;0	85	8;5	55	3;0
B202	4.71	4.08	6.00	3.00	4.71	4.08	6.00	3.00	92	91	8;1	65	5;1	118	13;7	65	5;3
B203	3.25	2.58	5.67	2.80	2.63	2.42	5.33	2.80	79	55	7;9	95	12;4	67	8;5	90	10;4
B204	4.38	3.54	5.38	5.67	1.88	1.46	3.13	3.83	94	55	8;0	55	3;0	96	14;1	55	4;8
B205	1.88	4.15	5.50	6.67	1.63	3.38	5.25	6.33	83	78	16;3	105	15;10	93	18;11	105	15;10
B206	4.00	4.92	5.17	8.00	3.63	3.75	4.67	8.00	83	78	11;3	75	7;1	108	18;5	75	7;1
B207	6.13	3.29	6.86	5.50	6.38	4.71	7.43	7.83	88	55	9;8	55	3;0	74	11;1	55	3;0

[a] Oral/Written comprehension/production evaluation range from 1 (very easy) to 9 (very difficult).
[b] Nonverbal IQ, Receptive vocabulary and Sentence Recall abilities are presented in standard scores (M = 100, SD = 15).
[c] Age-equivalent scores for the standardized tests. The CELF-5 interpretation manual does not provide specific chronological age equivalents below or above 3;00-15;10 but the following interpretations: ≤3;00 or ≥ 15;10 years old. Thus these two age points are used.

Appendix B. Sample items of the Catalan version of the Sentence Judgment task

Practice items

20. Grammatically correct-semantically correct sentences (GCSC)
La meva professora és molt bona i explica molt bé.
The my teacher is very good and explains very well.
'My teacher is very good and explains very well'.

21. Grammatically correct-semantic violation sentences (GCSV)
Viatgem en tortuga perquè és més ràpid que viatjar en cotxe.
Travel in turtle because is more quick than travel in car.
'We travel by turtle because is quicker than travelling by car'.

22. Grammar violation-semantically correct sentences (GVSC)
La Marta és al·lèrgica al préssec i no pot *lo tocar [tocar-lo].
The Marta is allergic to the peach and not can it touch.
'Marta is allergic to the peach and cannot *it touch [touch it]'.

23. Grammar violation-semantic violation (GVSV)
La noia *ros [rossa] va comprar l'última sabata a la floristeria.
The girl$_{FEM}$ blond$_{MASC}$ bought the last shoe in the florist.
'The blond[77] girl bought the last show in the florist'.

Experimental items

24. Grammatically correct-semantically correct sentences (GCSC)
Últimament el meu germà gran està dormint fins molt tard.
Lately the my brother old is sleeping to very late.
'Lately my older brother is sleeping until very late'.

25. Grammatically correct-semantic violation sentences (GCSV)
El Joan pensa que puc aparcar el llit al pàrquing.
The Joan thinks that can park the bed in the parking.
'Joan thinks that I can park the bed in the parking lot'.

26. Grammar violation-semantically correct sentences (GVSC)
Ahir l'Enric va comprar una televisió i *la ens va regalar [ens la va regalar].
Yesterday the Enric bought a television and it to us gifted.
'Yesterday Enric bought a television and gave *to us it [it to us] as a gift'.

27. Grammar violation-semantic violation (GVSV)
Avui *anat nosaltres hem [nosaltres hem anat] a l'aeroport per agafar el vaixell.
Today gone we have to the airport to take the boat.
'Today *gone we have [we have gone] to the airport to take the boat'.

[77] As in Spanish, adjectives in Catalan agree in gender and number with the noun. This example includes a gender violation (feminine noun and masculine adjective).

Appendix C. Sample items of the Catalan version of the Word-Length Judgment task

Practice items

Treballa-Tens ‘Work$_{3SG}$-Have$_{2SG}$’, Feu-Naveguen ‘Do/Make$_{2PL}$-Sail$_{3PL}$’

Experimental items

28. Congruent pairings (largest animal/object was the longest word)

Rinoceront-Pop or Helicòpter-Got

‘Rhinoceros-Octopus’ or ‘Helicopter-Glass’

29. Incongruent pairings (smallest animal/object was the longest word)

Canari-Porc or Bolígraf-Forn

‘Canary-pig’ or ‘Pen-Oven’

30. Big neutral pairings (big animals/objects independently of the word length)

Unicorn-Ruc or Armari-Llit

‘Unicorn-Donkey’ or ‘Closet-Bed’

31. Small neutral pairings (small animals/objects independently of the word length)

Esquirol-Talp or Forquilla-Flor

‘Squirrel-Mole’ or ‘Fork-Flower’