Article

Regular and Irregular Inflection in Different Groups of Bilingual Children and the Role of Verbal Short-Term and Verbal Working Memory

Elma Blom [1,2,*], Evelyn Bosma [1,3,4] and Wilbert Heeringa [4]

1   Department of Education and Pedagogy, Development and Education of Youth in Diverse Societies (DEEDS), Utrecht University, 3584 CS Utrecht, The Netherlands; e.bosma@hum.leidenuniv.nl
2   Department of Language and Culture, UiT The Arctic University of Norway, N-9037 Tromsø, Norway
3   Centre for Linguistics, Leiden University, 2300 RA Leiden, The Netherlands
4   Department of Linguistics, Fryske Akademy, 8911 DX Leeuwarden, The Netherlands; wheeringa@fryske-akademy.nl
*   Correspondence: w.b.t.blom@uu.nl

Abstract: Bilingual children often experience difficulties with inflectional morphology. The aim of this longitudinal study was to investigate how regularity of inflection in combination with verbal short-term and working memory (VSTM, VWM) influences bilingual children's performance. Data from 231 typically developing five- to eight-year-old children were analyzed: Dutch monolingual children ($N = 45$), Frisian-Dutch bilingual children ($N = 106$), Turkish-Dutch bilingual children ($N = 31$), Tarifit-Dutch bilingual children ($N = 38$) and Arabic-Dutch bilingual children ($N = 11$). Inflection was measured with an expressive morphology task. VSTM and VWM were measured with a Forward and Backward Digit Span task, respectively. The results showed that, overall, children performed more accurately at regular than irregular forms, with the smallest gap between regulars and irregulars for monolinguals. Furthermore, this gap was smaller for older children and children who scored better on a non-verbal intelligence measure. In bilingual children, higher accuracy at using (irregular) inflection was predicted by a smaller cross-linguistic distance, a larger amount of Dutch at home, and a higher level of parental education. Finally, children with better VSTM, but not VWM, were more accurate at using regular and irregular inflection.

Keywords: child bilingualism; inflectional morphology; nouns; verbs; plurals; participles; regularity; verbal short-term memory; verbal working memory

1. Introduction

Inflectional morphology is a locus of difficulty for bilingual children, but less is known about the role of regularity of inflection, and about the impact of verbal short-term and working memory on children's accuracy at using regular versus irregular inflection. Investigating these issues, the current study aimed to contribute to our understanding of how language-internal factors, child-internal factors, and their interactions determine language outcomes in bilingual children. Importantly, *the* bilingual child does not exist. Hoff (2013, p. 261) stated that "All children who are bilingual have in common that they have been exposed to two languages. Beyond that, the ways in which children's environments provide exposure to two (or more) languages vary enormously." This statement highlights the variation that exists within the bilingual population, which pertains to exposure and use (Francot et al. 2020). Others argue that a binary categorical distinction between monolinguals and bilinguals falls short in describing the actual variation that exists in both groups (Surrain and Luk 2017; Bialystok et al. 2012; Dixon et al. 2012; Kroll and Bialystok 2013). To explore variation among bilingual children and determine whether the observed patterns generalize across bilinguals, we distinguished in this study between bilinguals who are all learners of Dutch but who have either Arabic, Frisian, Tarifit or

Turkish as their other language. We examined bilingual speakers of these four languages in particular, because Frisian is the only officially recognized regional minority language in the Netherlands (Mercator European Research Centre on Multilingualism and Language Learning 2007) and Arabic, Tarifit and Turkish are the languages of two of the largest non-Western immigrant groups in the Netherlands (Den Ridder et al. 2020).

1.1. Inflection and Morphological Regularity

Various studies have shown that bilingual children make more inflection errors than their monolingual peers (Nicoladis et al. 2007; Rispens and De Bree 2015; Schwartz et al. 2009). That bilingual children perform less well on inflectional morphology is not surprising, given that their total language input is distributed across languages (Oller and Eilers 2002), which results in less input in one language compared to monolingual children learning that language. In addition, for many bilinguals, the age of onset to one of their languages is not from birth, resulting in a shorter length of exposure compared to same-age monolingual children.

Despite the unequal input to bilingual and monolingual children, it is not the case that bilingual children perform lower across-the-board. For example, investigating Hebrew-Dutch children, Rispens and De Bree (2015) found that bilingual and monolingual children performed similarly on Dutch regular past tense, but worse on Dutch irregular past tense. A similar observation was reported by Schwartz et al. (2009) who conclude that their monolingual and bilingual participants did not differ on regular plurals in Hebrew, while the bilinguals were less accurate than monolinguals on irregular forms. Relatively high accuracies with regular forms are not only found in between-child comparisons, but also emerge in within-child comparisons in studies that describe cross-linguistic patterns in bilingual French-English children (Nicoladis et al. 2007; Nicoladis and Paradis 2012; Paradis et al. 2011). These studies demonstrate higher accuracy for regular than irregular tense inflection with a larger difference between regulars and irregulars in English than French. In sum, regularity of inflection appears to be one of the key factors that modulates whether or not monolinguals outperform bilinguals on inflectional morphology.

Adopting a Usage-Based (UB) approach to language development, which presupposes that aspects of children's language input determine their language development, scholars have attributed the developmental differences between regular and irregular forms to frequency distributions in the input (e.g., Paradis et al. 2011; Rispens and De Bree 2015; Schwartz et al. 2009). In line with this approach, empirical studies have demonstrated that input distributions indeed predict bilingual children's accuracy at using regular and irregular inflection (Blom et al. 2012; Blom and Paradis 2013). The different acquisition rates of regular and irregular inflection originate, most likely, from relative type and token frequencies (Bybee 2007, 2008). For example, in English, regular past tense *-ed* has a high type frequency, which means that there are many different verbs that end in the past tense with *-ed*. The phonological variability of these verbs also tends to be large, which, together with a high type frequency, contributes to children's early productive use of *-ed*. Irregular past tense forms in English have generally a low type frequency and children's accuracy with irregular past tense is largely dependent on token frequency, which varies from verb to verb and results in different acquisition rates for individual verbs. Low token frequency irregular verbs will be late acquired, in particular by bilingual children as their input is distributed (Schwartz et al. 2009). The cross-linguistic differences between English and French may stem from the higher type and lower token frequencies of irregular forms in French compared to English (Nicoladis et al. 2007; Paradis et al. 2011).

Previous research on the acquisition of inflection in bilingual children has largely focused on regular and irregular verb inflection, specifically tense inflection in English. As part of the current study, we investigated noun plural and past participle formation in Dutch. To our knowledge, three studies have investigated this aspect of Dutch in monolingual and bilingual children (Boerma et al. 2017; Lalleman 1986; Verhoeven and Vermeer 1985). Lalleman (1986) and Verhoeven and Vermeer (1985) both found that the

monolinguals outperformed the bilinguals. Boerma et al. (2017) analyzed plurals and participles separately and used a two-wave longitudinal design. For plurals, they observed that the monolinguals were more accurate than the bilinguals regardless of wave and regularity. The monolingual children improved more than the bilingual children. This was related to the irregular noun forms that remained difficult for the bilinguals, in line with the suggestion by Schwartz et al. (2009). In the case of participles, a different pattern emerged. The monolinguals were more accurate than the bilinguals on regular and strong forms, but only a marginal difference was found for irregular participles, because these forms were difficult for bilingual and monolingual children alike. The current longitudinal study builds on these previous findings by investigating different bilingual groups, and asking how performance on regular and irregular noun plurals and past participles relates to child-internal resources such as verbal short-term memory and verbal working memory. We turn to these issues below.

1.2. Different Bilingual Groups and Cross-Language Distance

The bilingual children in the current study share the same majority language (Dutch), but have different minority languages (Arabic, Frisian, Tarifit, Turkish). In the context of this study, the notion 'minority language' is used as an umbrella for the languages spoken by a minority of the inhabitants in the country, and it contrasts with the notion 'majority language', which refers to the dominant language in the society that is spoken by the majority of the inhabitants (Hogan-Brun and Wolff 2003). In this sense, minority language encompasses heritage language, home language, native language, first language, or regional language, among other more specific notions.

Comparing performances across groups that have a different minority language can provide insight into the role of cross-language distance and transfer (Blom et al. 2020; Floccia et al. 2018). Several studies have investigated accuracy at using English tense marking inflection in English language learners in Canada dividing the sample of participating children into children learning a tense-marking inflecting minority language and children learning a non-tense marking isolating minority language (Paradis 2011; Blom et al. 2012). These studies consistently show that being familiar with tense inflection via the minority language is associated with fewer errors in using tense inflection in the majority language, pointing to effects of transfer. Separate analyses for English regular and irregular past tense inflection indicate that both are impacted by transfer effects (Blom and Paradis 2013). Other research shows that these transfer effects are persistent and that despite early and lengthy exposure to English, English tense inflection is a locus of errors if the minority (and native language) does not possess tense inflection (Paradis et al. 2016; McDonald 2000).

For the current study, we investigated children whose minority language is a migrant language (Arabic, Tarifit, Turkish) or a regional language (Frisian). All four languages are minority languages in the context of the Netherlands, where the study is situated, and they all have inflection. However, they overlap with Dutch, the majority language, in different degrees. Frisian is typologically close to Dutch, which is reflected in lexical and grammatical overlap, although there are also differences between Frisian and Dutch. In a previous study, we examined transfer effects in performance on Dutch noun plurals and past participles in the same Frisian-Dutch sample included in the current study by comparing accuracies on target forms that do and do not overlap between the two languages (Blom and Bosma 2016). Overlap was based on lexical and grammatical overlap, that is, if Dutch and Frisian have the same lexeme and similar suffixes and changes in the stem for plural or participle formation. The results revealed that the children performed better on overlapping items, confirming that cross-language proximity impacts on bilingual children's performance and leads not only to between-child variation but also to within-child variation. Arabic, Tarifit, and Turkish are all typologically distant from Dutch. Although all three languages have rich inflectional systems, we expected that the groups of children with Arabic, Tarifit or Turkish as the minority language would perform lower than the Frisian group on regular and irregular inflection in Dutch, because of much less phonological and morphological

overlap with Dutch. Arabic, Tarifit, and Turkish are quite similar to each other in terms of distance to Dutch (Blom et al. 2020). Yet, for the purpose of the current study, we did not collapse the data from the bilingual Arabic-Dutch, Tarifit-Dutch and Turkish-Dutch children, because previous research suggests notable differences in bilingual exposure, use and proficiency between these groups and patterns found for one group do not necessarily generalize to other groups (Scheele et al. 2010; Blom 2019).

1.3. Verbal Short-Term and Working Memory

According to the UB approach, children's accuracy at using inflection is in part determined by child-external input-related factors. Individual variation in children's ability to accurately use inflection is moreover expected to be impacted by child-internal resources, such as verbal short-term memory (VSTM) and verbal working memory (VWM). Several studies found that children with better VSTM and VWM also had higher language outcomes (Adams and Gathercole 1996, 2000; Baddeley et al. 1998; French and O'Brien 2008; Masoura and Gathercole 2005; Paradis 2011; Service 1992; Verhagen and Leseman 2016), suggesting that language knowledge in long-term memory is associated with children's ability to *temporarily* store and manipulate information. It is still largely an open question whether VSTM and VWM have differential effects, and whether regular and irregular aspects of language show different relations with VSTM and VWM.

According to Baddeley's (2003) well-known working memory model, VSTM consists of two subcomponents: one to hold memory traces in mind and another to prevent these traces from decay through subvocal rehearsal. As such, VSTM is thought to facilitate language acquisition in native and child second language (L2) learners. Recent studies with bilingual children support this hypothesis. Paradis (2011) found that VSTM predicts receptive vocabulary and the production of tense inflection in English language learners who are between ages 4 and 7 years. Another study found that VSTM predicts receptive vocabulary and grammar in 5- and 6-year-old monolingual Dutch and bilingual Turkish-Dutch children (Verhagen and Leseman 2016). In this latter study, the same relations emerged for monolinguals and bilinguals, and for grammar skills at the sentence level and at the word level. At the word level, the same (standardized) instrument was used as in the current study, targeting children's ability to produce correct noun plurals and past participles. Verhagen and Leseman (2016) did, however, not distinguish the regular and irregular items that were both included in this task. It is therefore unknown whether or not VSTM differently related to regular and irregular inflection.

VWM refers to the ability to manipulate temporarily stored information, and involves storage and concurrent processing of information. VWM is, for example, measured with a listening span task in which children need to comprehend a sentence and provide a truth-value judgment about this sentence, while, at the same time, remember a digit presented auditorily after the last word of each sentence or via a digit span task in which children are asked to repeat sequences of auditorily presented digits in reversed order. VWM is thus involved in more complex mental processes, which is confirmed in research with children showing that VWM performance is constrained both by VSTM capacity and executive processes such as domain-general attention (Magimairaj and Montgomery 2012). Two studies observed significant relations between VWM and second language (L2) grammar outcomes in 15- to 16-year-old youth (Kormos and Sáfár 2008) and 7- to 8-year-old children (Engel de Abreu and Gathercole 2012). These studies investigated L2 learning in instructional settings, and did not focus on inflectional morphology. Verhagen and Leseman's (2016) study demonstrates that VWM also predicts grammar, including accurate use of noun plurals and past participles, in a naturalistic learning setting. It is, however, unknown whether this effect was the same for regular and irregular items.

There are reasons to hypothesize that VWM has a differential effect on regular and irregular inflection. Investigating native-speaking Arabic children in Israel who were, on average, 11 years old, Cohen-Mimran et al. (2013) observed that VWM was only related to regular inflection. In their study, children participated in a listening span task in which they

were asked to recall target words at the end of sentences. After hearing a series of sentences, they had to recall the target words which could be regular or irregular. Recall of words with regular morphology was found to be significantly poorer than recall of words with irregular morphology. The authors argue that the decomposition of regular forms imposes extra load and draws on executive processes, explaining the differential effects for regulars and irregulars. A similar explanation is given by Bosma et al. (2017) who found that VWM, but not VSTM, predicted Frisian-Dutch bilingual children's ability to detect cross-linguistic phonological regularities. A recent study with Turkish adults found no effects of VWM on the processing of English L2 regular past tense morphology (Rizaoğlu and Gürel 2020). According to the authors, the absence of a significant outcome may, however, be related to high VWM scores and a lack of variation.

1.4. The Present Study

The aim of this longitudinal study is to investigate bilingual children's performance on noun plurals and past participles in Dutch focusing on effects of regularity, cross-language distance, and the role of VSTM and VWM. Previous research investigated effects of regularity, but it is unknown whether findings are the same across bilingual groups that vary in cross-language distance, and if effects of VSTM and VWM interact with regularity. Furthermore, the longitudinal design made it possible to investigate developmental patterns.

First, we hypothesized that owing to the distributed input of bilingual children, bilingualism would affect children's accuracy at using inflection and that this effect would be moderated by regularity and cross-language distance. The following predictions were formulated. We expected that the bilingual children in our study would be outperformed by the monolingual children (Nicoladis et al. 2007; Lalleman 1986; Verhoeven and Vermeer 1985; Boerma et al. 2017), and that this effect would be more pronounced for irregulars than regulars (Rispens and De Bree 2015; Schwartz et al. 2009; Boerma et al. 2017). It was predicted that the gap between bilinguals and monolinguals would be larger for the Arabic-Dutch, Tarifit-Dutch or Turkish-Dutch children than for Frisian-Dutch children. The reason is that positive transfer is facilitated if the cross-language distance is smaller (Blom and Bosma 2016; Floccia et al. 2018). Second, we hypothesized that regular and irregular inflection would differ in accuracy because of differences between regular and irregular inflection in input distributions, specifically, differences in type-token frequency balance. We predicted that children would perform more accurately using regular inflection compared to irregular inflection, in line with previous findings for English and French (Nicoladis et al. 2007; Nicoladis and Paradis 2012; Paradis et al. 2011). Rispens and De Bree (2015) suggest, moreover, that age moderates the differences in accuracy between regulars and irregulars as an effect of accumulating input. This may indeed be a viable hypothesis, as the older children are, the more input they have accumulated and the more likely it is that the specific effects of input distributions are not noticeable anymore as soon as input thresholds for mastery have been reached. Third, we hypothesized that children's accuracy at using inflection would be dependent on cognitive resources, specifically VSTM and VWM capacity, with differential effects for regular and irregular inflection. We predicted that VSTM and VWM would be positively related to accuracy at using inflection (Verhagen and Leseman 2016), regardless of bilingual group, and that better VWM scores would predict higher accuracy at using regular but not irregular inflection (Bosma et al. 2017; Cohen-Mimran et al. 2013).

2. Materials and Methods

2.1. Participants

For this longitudinal study, data from 231 children were analyzed. All children were typically developing with no indication of language and speech disorders, as judged by the Questionnaire for Parents of Bilingual Children (PaBiQ; Tuller 2015). They were assigned to five different groups: monolingual children ($N = 45$), Frisian-Dutch bilingual children ($N = 106$), Turkish-Dutch bilingual children ($N = 31$), Tarifit-Dutch bilingual

children (*N* = 38) and Arabic-Dutch bilingual children (*N* = 11). The children were followed longitudinally, with three measurements over one-year intervals. All children were 5 or 6 years old at time 1, 6 or 7 years old at time 2, and 7 or 8 years old at time 3. Bilingual children were included in the dataset if their exposure to the non-Dutch language at home was more than 20% of the time. The groups are of unequal sizes because we combined two separate datasets for the current study, one focusing on bilingualism in migration contexts (Turkish-Dutch, Tarifit-Dutch) and another focusing on bilingualism in the context of regional minority languages (Frisian-Dutch). Only during testing did it turn out that some children of Moroccan descent spoke Arabic instead of Tarifit. We included these children, treating them as a separate bilingual group. All children were born in the Netherlands, and exposed to Dutch and the minority language from birth in different degrees with considerable variation in bilingual exposure in all groups. Participant characteristics are given in Table 1. As age, parental education and non-verbal IQ have previously been found to influence the acquisition of inflectional morphology (Paradis 2011), these were included as control variables in the present study.

Table 1. Means (standard deviations) for age in months at time 1, 2 and 3, non-verbal intelligence (WNV (Wechsler Non-verbal Scale of Ability), standardized norm score), parental education (EdPar, measured on a 9-point scale) and current use of Dutch at home (Use Dutch, in percentages), and gender distribution in the three groups.

Characteristic	Monolinguals (*N* = 45)	Bilinguals Frisian (*N* = 106)	Bilinguals Turkish (*N* = 31)	Bilinguals Tarifit (*N* = 38)	Bilinguals Arabic (*N* = 11)
Age time 1	71 (8)	71 (7)	71 (7)	66 (7)	72 (5)
Age time 2	82 (8)	82 (7)	79 (9)	77 (7)	84 (5)
Age time 3	94 (8)	94 (7)	91 (9)	89 (6)	96 (5)
WNV	107 (15)	107 (15)	103 (13)	94 (14)	95 (14)
EdPar	6.59 (1.93)	6.84 (1.23)	4.85 (2.31)	5.06 (2.30)	4.80 (2.26)
Use Dutch	-	30 (22)	45 (12)	59 (10)	52 (16)
Girls/boys	20/25	50/56	15/16	20/18	8/3

Non-verbal IQ was measured with the short version of the Wechsler Non-verbal Scale of Ability (WNV; Wechsler and Naglieri 2006). This comprised the subtest Matrices, in which children had to complete figural matrices by selecting the missing piece from four or five response options, and the subtest Recognition, in which children looked at geometric designs for three seconds and then had to select the viewed stimulus from four or five response options. Parental education was calculated as the average educational level of both parents, based on the PaBiQ (Tuller 2015). Educational level was defined as the highest degree obtained on a 9-point scale, ranging from no education (1) to a university degree (9). The PaBiQ also provided information about the current use of the Dutch language at home, which was measured relative to the total amount of language input that the child received from his/her mother, father, siblings and other adults who looked after the child at least once per week. For each of these people, the question had to be answered how often (s)he spoke the Dutch language to the child: 'never' (0%), 'seldom' (25%), 'sometimes' (50%), 'usually' (75%) and 'always' (100%).

2.2. Measures

2.2.1. Inflectional Morphology

Inflectional morphology was measured with the Word Formation subtest which is part of the Taaltest Alle Kinderen (TAK, "Language Assessment for All Children", Verhoeven and Vermeer 2001). The Word Formation test includes 12 noun items testing pluralization, and 12 verb items testing participle formation. Plurals or participles that did not undergo a change to the stem vowel were classified as regular (8 nouns, 4 verbs)

whereas plurals and participles that showed a stem vowel change were classified as irregular (4 nouns, 8 verbs). The items are listed in Appendix A.

Items were scored as incorrect if there was a morphological error related to the specific properties of the inflected form. For noun pluralization, this included the omission of the plural suffix, the use of an incorrect plural suffix, and no lengthening of the stem vowel in cases where this was required, as in *weg-wegen* (ʋɛx-'ʋe:gə) 'road-roads'. Phonological errors in the stem but not in the target morpheme, e.g., *kranden* instead of *kranten* 'newspapers', were not scored as incorrect. Final *-n* deletion was also not considered a mistake, since this is not uncommon in colloquial Dutch (Booij 1995, p. 141). For participle formation, items were scored as incorrect when the prefix or suffix was omitted or used incorrectly, or when there was an error with the stem, i.e., no or incorrect changes to the stem.

2.2.2. Verbal Short-Term and Working Memory

Verbal short-term memory and verbal working memory were measured with a Forward and a Backward Digit Span task, respectively. The tasks were based on the Alloway Working Memory Assessment (AWMA; Alloway 2012) and were translated to Dutch. The children heard sequences of digits, which they had to repeat in the same (Forward Digit Span) or in reverse order (Backward Digit Span). The Forward Digit Span is considered a measure of verbal short-term memory, as the digits only need to be stored, whereas the Backward Digit Span is considered a measure of verbal working memory, as the digits need to be stored and processed to be able to repeat them in reverse order (Alloway et al. 2008).

The AWMA procedure was applied for scoring (Alloway 2012). Both the Forward and the Backward Digit Span task consisted of seven blocks of six trials, and for each correct trial, the child received one point. This means that there was a maximum of six points per block and a maximum of 42 points per task. Trials were scored as incorrect if the sequence was (partly) incorrect, if children omitted one or more digits, or if they recalled one or more digits incorrectly. Each task started with a block with sequences of one digit, after which the length of the sequences increased with one digit for each subsequent block. The child automatically continued with the next block, when (s)he repeated the first four, or four out of the first five trials, within one block correctly, in which case (s)he received a score of 6 or 5, respectively. In other words, the child automatically received one point for each skipped trial. When the child answered three trials within one block incorrectly, the task stopped immediately.

2.3. Procedure

This research was screened by the Standing Ethical Assessment Committee of the Faculty of Social and Behavioral Sciences at Utrecht University. Criteria were met and further verification was not deemed necessary. Participants were recruited through primary schools in the Netherlands, which distributed consent forms and information folders about the experiment among the parents of the children. In accordance with the Declaration of Helsinki, the parents of the participating children all signed a consent form. All children were tested individually in a quiet room at school or at home. They were tested by experimenters who had a native command of Dutch and, in the case of the bilinguals, also of Frisian, Turkish, Tarifit or Arabic. The tasks in the present study were part of a larger test battery that included language, working memory and attention tasks. For all of the measures reported in this study, the language of instruction was Dutch.

2.4. Data Analysis

Our aim was to investigate monolingual and bilingual children's accuracy at using regular versus irregular inflection, and specifically to determine whether (1) different groups of bilingual children score lower on irregular inflection, but not on regular inflection, compared to monolingual children, (2) differential outcomes for regular and irregular inflection are modulated by age, (3) VSTM and VWM predict outcomes on both regular and irregular inflection. Children's responses on the 12 noun and 12 participle items

were the binary outcome (correct, incorrect). The main factors of interest were: Group (monolinguals, Arabic-Dutch, Frisian-Dutch, Tarifit-Dutch, and Turkish-Dutch bilinguals), Regular (irregular, regular), Age, VSTM and VWM.

2.5. Preliminary Analyses

Preliminary analyses were carried out to examine whether the groups showed differences on any of the child-level predictor variables that were included in the main analyses. We first examined differences in Age, VSTM and VWM between the five groups. For each of these variables, we performed a linear mixed-effects model with the respective variables as dependent variables, Group as fixed factor, and Subject and Time (1, 2, 3) as random intercepts. For Age, Subject prevented the model from converging, therefore it was left out as random intercept. We used the R function lmer from the lme4 package. Pairwise comparisons were carried out with the R function pairs and emmeans from the emmeans package using the Šidàk correction. Next, the variables Parental education, Non-verbal intelligence and Use Dutch (the percentage of current use of Dutch) in the homes of the bilingual children were compared across the five groups. Since these variables were only measured at time 1, there is just one measurement per subject. Therefore, no random intercepts for Subject and Time were needed. Since a classical ANOVA did not meet the assumptions, we used the R function pairwise.wilcox.test with Holm correction. Finally, using Fisher's Exact test we compared the proportions females/males between the groups using the R function fisher.test.

The Tarifit-Dutch bilinguals were significantly younger than the monolinguals ($t = -5.54$, $p < 0.001$), Arabic-Dutch ($t = -4.83$, $p < 0.001$), Frisian-Dutch ($t = -6.23$, $p < 0.001$) and Turkish-Dutch bilinguals ($t = -3.66$, $p < 0.001$). The Tarifit-Dutch bilinguals scored significantly lower on VSTM than the monolinguals ($t = -4.24$, $p < 0.001$) and the Frisian-Dutch bilinguals ($t = -3.44$, $p < 0.01$). Their VWM scores were lower than those of the Frisian-Dutch bilinguals ($t = -3.46$, $p < 0.01$). The parents of the monolinguals were significantly more highly educated than the parents in the Tarifit-Dutch ($p < .05$) and Turkish-Dutch bilingual groups ($p < 0.05$), and the parents of the Frisian-Dutch bilinguals were significantly more highly educated than the parents in the Arabic-Dutch ($p < 0.05$), Tarifit-Dutch ($p < 0.001$) and Turkish-Dutch groups ($p < 0.01$). The Tarifit-Dutch bilinguals had significantly lower non-verbal intelligence scores than the monolinguals ($p < 0.01$), Frisian-Dutch ($p < 0.001$) and Turkish-Dutch bilinguals ($p < 0.05$). In the homes of the Arabic-Dutch, Tarifit-Dutch and Turkish-Dutch bilinguals, significantly more Dutch was used with the children compared to the Frisian-Dutch bilinguals (Arabic: $p < 0.01$, Tarifit: $p < 0.001$, Turkish: $p < 0.01$), and in the Tarifit-Dutch homes, significantly more Dutch was used with the children than in the Turkish-Dutch homes ($p < 0.001$). There were no significant differences in the proportions of boys and girls for language group pairs.

2.6. Main Analyses

To test whether our predictions were borne out by the data and pinpoint significant explanatory factors, we analyzed the data using mixed-effects models: one model on the basis of all children, and a second model on the basis of only the bilingual subjects. We ran this second model to look more closely into differences between the four bilingual groups by including the role of the variable Use Dutch. This variable could not be included in the first model with the monolinguals because the monolingual children did not vary on the current use of Dutch at home.

In both models, in addition to the variables of interest (Group, Regular, Age, VSTM, VWM), the variables Sex (boy, girl), Parental education, Non-verbal intelligence, Word type (noun, verb), and Lemma frequency (based on the Corpus Spoken Dutch, Oostdijk 2000) were included because these may predict additional variance, allowing us to better understand the effects of the factors of interest. Correlations between the fixed-effects predictors are in Appendix B.

For the first model, we started by entering the following variables in the model: Age, Sex, Group (monolingual, Frisian-Dutch, Arabic-Dutch, Tarifit-Dutch and Turkish-Dutch bilingual), Parental education, Non-verbal intelligence, VSTM, VWM, Word type, Regular, and Lemma frequency. Additionally, as interactions, we added all possible combinations of two variables. Since the variables Sex, Word type and Lemma frequency were not found to yield significant main effects, we did not include interactions with these variables. In order to run the model, we used the function glmer of the R Package lme4 (Bates et al. 2015). Subject and Word were entered as random intercepts.

Given the binary response variable, we ran a logistic model with model link function 'logit' and optimizer 'nlminb' from the R package optimx (Nash and Varadhan 2011). In our experience, the optimizer 'nlminb' makes models converging where other optimizers do not. Nevertheless, there were four interactions that prevented the model to converge: Age:Non-verbal intelligence, Age:Regular, Parental education:Regular, VSTM:VWM. We removed those interactions. Given the large number of main effects and interactions, it is tempting to use automatic variable selection procedures such as stepwise forward or backward selection. However, these methods have strongly been criticized (Flom and Cassell 2007). Among other problems, models resulting from a stepwise selection produce *p*-values that are biased towards 0. Therefore, we decided to continue with the full model as was suggested as one of the alternatives by Flom and Cassell (2007). By using the function Anova from the car package for each of the fixed-effects, Wald χ^2 tests were calculated and *p*-values obtained (Fox and Weisberg 2019). We chose the type III test because of the presence of (significant) interactions (see: https://www.r-bloggers.com/2011/03/anova-%E2%80%93-type-iiiiii-ss-explained/, accessed on 19 March 2021).

After running the full model, the effects of significant fixed-effects were examined. Pairwise comparisons were carried out with the R function pairs and emmeans from the emmeans package, which we chose for the more conservative Šidàk correction. This correction is slightly less conservative than the Bonferroni correction, but more powerful (Bretz et al. 2016). Graphs were plotted with functions from the ggeffects package which uses the predictions generated by the model when one holds the non-focal variable(s) constant and varies the focal variable(s) (Lüdecke 2018). For plotting (interaction) graphs of only categorical variables, we used ggemmeans, and for plotting continuous variables or interactions that include a continuous variable, we used the function ggpredict. Confidence intervals were based on the standard errors of the predicted values, assuming normal distribution ($\pm$1.96 * SE).

For the second model, the same procedure was followed as for the complete data set (see above). The same variables were entered in the model, and the variable Use Dutch was added. As interactions, we added all possible combinations of two variables. Since the variables Sex, Word type and Lemma frequency were not found to yield significant main effects, we did not include interactions with these variables. We entered Subject and Word as random intercepts. Twelve interactions prevented the model to converge, and were removed: Age:VSTM, Age:VWM, Group:Non-verbal intelligence, Group:VWM, Non-verbal intelligence:VWM, VSTM:Regular, VWM:Regular, Use Dutch:Age, Use Dutch:Parental education, Use Dutch:Non-verbal intelligence, Use Dutch:VSTM, Use Dutch:VWM.

3. Results

Children's accuracies at regular and irregular inflection are displayed in Table 2, distinguishing between regular and irregular inflection. VSTM and VWM scores are shown in Table 3.

Table 2. Unweighted grand means and standard deviations (based on the means and standard deviations of the proportions of correctly used regular and irregular inflection of the three measurement times). For groups where the number of participants differs per time, the number of participants is given for each time (time 1/time 2/time 3).

		Regular		Irregular	
Group	**Number of Children**	**Mean**	**Standard Deviation**	**Mean**	**Standard Deviation**
Monolinguals	44/45/44	0.93	0.25	0.66	0.46
Frisian-Dutch	106	0.93	0.25	0.55	0.48
Turkish-Dutch	20/24/21	0.73	0.44	0.32	0.45
Tarifit-Dutch	29/32/32	0.82	0.38	0.43	0.47
Arabic-Dutch	11	0.88	0.32	0.53	0.48

Table 3. Unweighted grand means and standard deviations (based on the means and standard deviations of the proportions of correctly used regular and irregular inflection of the three measurement times). For groups where the number of participants differs per time, the number of participants is given for each time (time 1/time 2/time 3).

		VSTM		VWM	
Group	**Number of Children**	**Mean**	**Standard Deviation**	**Mean**	**Standard Deviation**
Monolinguals	44/45/44	23.24	4.11	14.61	3.14
Frisian-Dutch	106	22.25	3.62	14.76	3.12
Turkish-Dutch	20/24/21	21.15	3.43	13.34	3.00
Tarifit-Dutch	29/32/32	19.96	3.05	13.10	2.69
Arabic-Dutch	11	22.36	3.17	13.91	2.69

Note: VSTM = verbal short-term memory, VWM = verbal working memory.

3.1. Mixed-Effects Modeling: All Children

In the first model, we analyzed the data of all children, both monolingual and bilingual. The results are presented in Table 4. In this section, we continue by discussing the significant fixed-effects one by one.

Table 4. Results of the mixed-effects model analysis of all children.

	χ^2	Df	Sig.
(Intercept)	1.05	1	n.s.
Age	78.95	1	$p < 0.001$
Sex	0.65	1	n.s.
Group	35.46	4	$p < 0.001$
EdPar	1.76	1	n.s.
WNV	12.94	1	$p < 0.001$
VSTM	20.94	1	$p < 0.001$
VWM	1.14	1	n.s.
Word type	0.09	1	n.s.
Regular	26.20	1	$p < 0.001$
Lemma frequency	3.54	1	n.s.
Age:Group	6.94	4	n.s.
Age:EdPar	1.36	1	n.s.
Age:VSTM	0.66	1	n.s.
Age:VWM	0.02	1	n.s.
Group:EdPar	10.30	4	$p < 0.05$
Group:WNV	6.26	4	n.s.
Group:VSTM	13.20	4	$p < 0.05$

Table 4. *Cont.*

	χ^2	Df	Sig.
Group:VWM	3.12	4	n.s.
Group:Regular	46.66	4	$p < 0.001$
EdPar:WNV	0.39	1	n.s.
EdPar:VSTM	1.69	1	n.s.
EdPar:VWM	1.69	1	n.s.
WNV:VSTM	3.11	1	n.s.
WNV:VWM	2.53	1	n.s.
WNV:Regular	7.69	1	$p < 0.01$
VSTM:Regular	1.41	1	n.s.
VWM:Regular	2.18	1	n.s.

Note: EdPar = parental education, WNV = non-verbal intelligence, VSTM = verbal short-term memory, VWM = verbal working memory.

As to the main effects, we found that higher age, higher non-verbal IQ scores (WNV) and higher VSTM significantly improved children's accuracy. Since the variables Group and Regular are involved in interactions, the results for the individual variables may be misleading. Therefore, we focus on the four significant interactions, starting with the two interactions that pertain to the predictions we formulated (Group:Regular, Group:VSTM). The figures that pertain to interactions that are not of direct relevance to our hypotheses can be found in Appendix C.

The interaction between Group and Regular is shown in Figure 1. Children performed more accurately at regular forms compared to irregular forms. For all five groups, this difference was significant (monolinguals: $z = -5.12$, $p < 0.001$; Frisian-Dutch bilinguals: $z = -6.29$, $p < 0.001$; Arabic-Dutch bilinguals: $z = -4.70$, $p < 0.001$; Tarifit-Dutch bilinguals: $z = -4.64$, $p < 0.001$; Turkish-Dutch bilinguals: $z = -5.50$, $p < 0.001$). Figure 1 shows that the difference between regulars and irregulars was smallest for monolinguals, and largest for Turkish-Dutch bilinguals. The difference between the groups was smaller for regular inflection than for irregular inflection, and for regular inflection, the 95% confidence intervals were smaller than for irregular inflection. For regular inflection, we found that the Turkish-Dutch bilinguals were less accurate than the monolinguals ($z = 4.59$, $p < 0.001$) and the Frisian-Dutch bilinguals ($z = 4.49$, $p < 0.001$). For irregular inflection, it turned out that the monolinguals were more accurate than the Frisian-Dutch bilinguals ($z = 3.96$, $p < 0.01$) and Turkish-Dutch bilinguals ($z = 5.66$, $p < 0.001$), and that the Turkish-Dutch bilinguals were less accurate than the Tarifit-Dutch bilinguals ($z = 3.50$, $p < 0.05$) and Frisian-Dutch bilinguals ($z = 3.28$, $p < 0.05$).
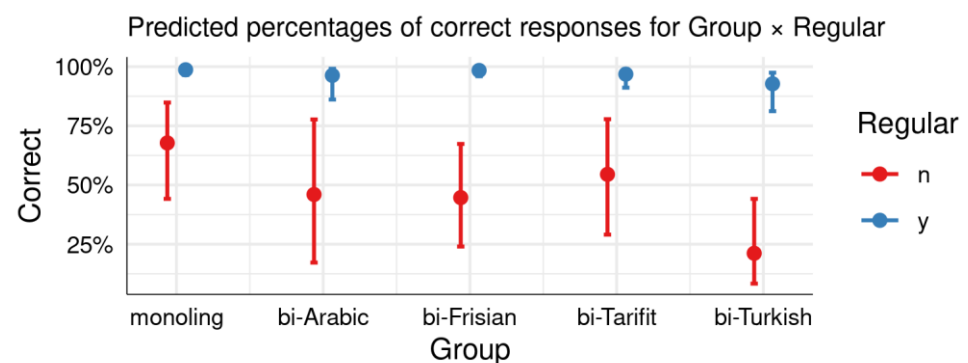

Figure 1. Interaction between group and regularity of inflection.

Figure 2 shows the interaction between Group and VSTM. The better children scored on VSTM, the more accurately they used inflection, except for the Turkish-Dutch group.
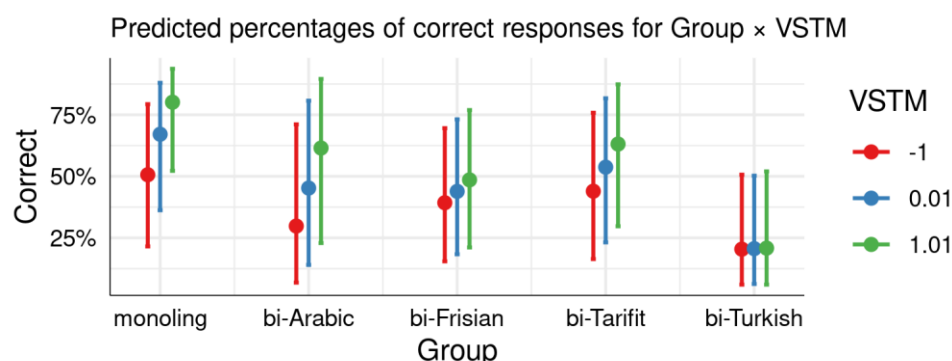
Figure 2. Interaction between group and verbal short-term memory (VSTM).

Two other interactions were significant: Group:Parental education (Figure A1) and Non-verbal intelligence:Regular (Figure A2). For each of the groups, we found that higher parental education was associated with higher accuracy at using inflection, but for the Turkish-Dutch and Frisian-Dutch bilingual groups, the effect of parental education was stronger than for the other groups. The interaction between non-verbal intelligence and regularity shows that with increasing intelligence scores, the difference between accuracy at regular and irregular forms became smaller. Figure A2 also suggests that the effect of non-verbal intelligence was larger for irregular inflection than for regular inflection. The 95% confidence intervals were larger for irregular inflection than for regular inflection.

3.2. Mixed-Effects Modelling: Bilingual Children

In the second model, we analyzed the data of the bilingual children only, so that we could also examine the effect of the current use of Dutch at home (Use Dutch). The results are presented in Table 5. We continue by discussing the significant fixed-effects one by one.

Table 5. Results of the mixed-effects model analysis of only bilinguals.

	χ^2	Df	Sig.
(Intercept)	0.03	1	n.s.
Age	275.84	1	$p < 0.001$
Sex	3.01	1	n.s.
Group	31.89	3	$p < 0.001$
EdPar	1.90	1	n.s.
Use Dutch	51.09	1	$p < 0.001$
WNV	7.04	1	$p < 0.01$
VSTM	6.95	1	$p < 0.01$
VWM	2.62	1	n.s.
Word type	0.28	1	n.s.
Regular	29.19	1	$p < 0.001$
Lemma frequency	3.33	1	n.s.
Age:Group	5.29	3	n.s.
Age:EdPar	0.74	1	n.s.
Age:WNV	1.38	1	n.s.
Age:Regular	29.81	1	$p < 0.001$
Group:EdPar	8.24	3	$p < 0.05$
Group:VSTM	6.04	3	n.s.
Group:Regular	4.32	3	n.s.
EdPar:WNV	0.02	1	n.s.
EdPar:VSTM	4.26	1	$p < 0.05$
EdPar:VWM	5.43	1	$p < 0.05$
EdPar:Regular	0.47	1	n.s.
Use Dutch:Group	5.06	3	n.s.
Use Dutch:Regular	71.82	1	$p < 0.001$

Table 5. *Cont.*

	χ²	Df	Sig.
WNV:VSTM	2.44	1	n.s.
WNV:Regular	4.97	1	$p < 0.05$
VSTM:VWM	0.13	1	n.s.

Note: EdPar = parental education, WNV = non-verbal intelligence, VSTM = verbal short-term memory, VWM = verbal working memory.

As to the main effects, we found that higher age, non-verbal intelligence (WNV) scores and VSTM scores significantly improved children's accuracy at using inflection. More frequent use of Dutch at home was also associated with higher inflection accuracy. Pairwise comparisons indicated that fewer errors were made with regular than irregular inflection ($p < 0.001$) and that the Frisian-Dutch and the Arabic-Dutch group performed more accurately than the Turkish-Dutch group (respectively $p < 0.001$ and $p < 0.05$). Although we are warned that the results for Group and Regular may be misleading since these variables are involved in interactions, they were not contradicted by the significant interactions in which they are involved. The significant interactions are discussed below, starting with those that involve Regular (Age:Regular, Use Dutch:Regular, WNV:Regular), followed by the interactions that involve Group (Group:Parental education). Lastly, the interactions with VSTM and VWM are presented (Parental education:VSTM, Parental education:VWM). The figures that illustrate interactions that are not of direct relevance to our hypotheses can be found in Appendix D.

The interactions between children's age and regularity, use of Dutch at home and regularity, and children's non-verbal intelligence scores are plotted in Figure 3 and in Figures A3 and A4 in Appendix D. The figures show that with increasing age, use of Dutch, and non-verbal intelligence scores, the difference between the scores for regulars and irregulars became smaller. Additionally, the figures suggest that the effects of age, use of Dutch, and non-verbal intelligence were larger for irregular inflection than for regular inflection. The 95% confidence intervals were larger for irregular inflection than for regular inflection.

Figure 3. Interaction between age and regularity of inflection.

The interaction between group and parental education is illustrated in Figure A5. For each of the groups, higher parental education was associated with higher accuracy, but for Turkish-Dutch and Frisian-Dutch groups, the effect of parental education was stronger than for the other groups.

Figures A6 and A7 illustrate the interactions between parental education and, respectively, VSTM and VWM. With increasing level of parental education and with better VSTM scores, children's accuracy on inflection improved. The positive effect of VSTM disappeared when parental level of education increased. Figure A7 shows that with increasing level of parental education, children's inflection accuracy improved. However, for low

values of parental education, lower inflection accuracy was associated with higher VWM scores, whereas for high values of parental education, it was the other way around.

4. Discussion

Inflectional morphology can be a locus of errors in bilingual children. The aim of this longitudinal study was to examine how this effect is influenced by regularity of inflection, determine the role of verbal short-term and working memory (VSTM, VWM), and to investigate interactions between these language- and child-level variables. In addition, unlike most previous research, this study was focused on Dutch and, importantly, included four groups of typically developing five- to eight-year-old bilingual children (Arabic-Dutch, Frisian-Dutch, Tarifit-Dutch and Turkish-Dutch), allowing us to explore variation within the heterogeneous population of bilingual children. VSTM and VWM were measured with a Forward and a Backward Digit Span task, respectively. A mixed-effects modeling approach was used in which we included not only the target variables (regularity, group, VSTM/VWM) as predictors of children's accuracy at using inflection, but also additional variables at the language-level (lemma frequency, word type) and child-level (age, sex, non-verbal intelligence, parental education, use of Dutch at home), enabling a better and more nuanced understanding of the impact of the target variables.

4.1. Regularity

Previous research on French-English bilinguals has shown that children's performance on regular inflection is more accurate than performance on irregular inflection, and that the gap between regular and irregular is larger in English than in French (Nicoladis et al. 2007; Nicoladis and Paradis 2012; Paradis et al. 2011). This difference between acquisition rates of regular and irregular forms is probably related to unequal input distributions and the relatively low type frequency of irregular forms, in line with a UB approach to language acquisition (Bybee 2007, 2008). The results of the current study indicate that higher accuracies for regular than irregular inflection are also found for children learning Dutch. This effect was found regardless of lemma frequency and whether items were nouns or verbs.

The difference between regular and irregular inflection was significant in all five groups in this study, but the size of the effect varied across groups. The Turkish-Dutch bilinguals showed the largest difference between regular and irregular inflection, whereas the difference was the smallest for monolinguals (in the analyses with all children). Interestingly, in the analysis with only bilingual children, we found a significant interaction between regularity and use of Dutch instead of an interaction between group and regularity. More use of Dutch at home was associated with a smaller difference between accuracy at using regular and irregular inflection. Use of Dutch was not included in the model with all children, as, due to how this variable was measured, the monolingual children did not vary in this respect. However, it is likely that the small difference between regular and irregular inflection found for the monolinguals stems from frequent exposure to and use of Dutch in this group.

In addition to group, two other factors interacted with regularity. Both analyses revealed that with increasing non-verbal intelligence, the gap between regular and irregular inflection narrowed. Regardless of their non-verbal intelligence scores, children were able to use regular inflection accurately, whereas irregular inflection was more difficult for children with lower non-verbal intelligence scores than for their peers with higher non-verbal intelligence scores. The effect of non-verbal intelligence replicates previous research showing that non-verbal intelligence predicts bilingual children's correct production of verb morphology (Paradis 2011), and demonstrates that analytical reasoning is a component of language aptitude in children. A second interaction effect was found for age. This interaction was only part of the statistical model with bilinguals, as the same model with all children did not converge. With older age, the differences in accuracy between regular and irregular inflection became smaller, in line with Rispens and De Bree (2015). This effect can

be attributed to increasing accuracy at using irregular inflection, as regular inflection was highly accurate throughout the age range investigated.

To conclude, between-group differences are more likely for irregular than regular forms. The effect of regularity is caused by the relative difficulty of irregular forms. More use of Dutch at home, an older age, and better learning potential/analytical reasoning predict fewer errors with irregulars and are associated with a smaller gap between regular and irregular inflection.

4.2. Group

Previous research has shown that monolingual children make fewer inflection errors than bilingual children (Nicoladis et al. 2007; Lalleman 1986; Verhoeven and Vermeer 1985; Boerma et al. 2017), in particular if accuracy at using irregular forms is compared across the two groups (Rispens and De Bree 2015; Schwartz et al. 2009; Boerma et al. 2017). We found effects of group, which were more pronounced for irregulars than regulars, but the effects of group did not conform to a binary distinction between monolinguals and bilinguals, and rather display a distinction between monolinguals versus Arabic-Dutch/Frisian-Dutch/Tarifit-Dutch bilinguals versus Turkish-Dutch bilinguals. For example, using regular inflection, Turkish-Dutch bilingual children made more errors than monolinguals, but no significant differences emerged between Tarifit-Dutch or Frisian-Dutch bilingual children on the one hand and monolingual children on the other hand. Additionally, monolingual children were more accurate than Frisian-Dutch bilinguals and Turkish-Dutch bilinguals at using irregular inflection, but the difference between monolinguals and Tarifit-Dutch bilinguals did not reach significance. We turn to this outcome below. The findings confirm the heterogeneity in the bilingual population which is also emphasized by other researchers (Surrain and Luk 2017; Bialystok et al. 2012; Dixon et al. 2012; Kroll and Bialystok 2013).

Our predictions about cross-linguistic distance were only partially borne out. Cross-linguistic distance is the smallest for the Frisian-Dutch group, hence this group was expected to be the bilingual group with the highest accuracy. The Frisian-Dutch group indeed performed significantly better than the Turkish-Dutch group on both regular and irregular inflection, in line with our predictions. However, the Frisian-Dutch group did not perform more accurately than the Arabic-Dutch and Tarifit-Dutch groups. Moreover, whereas the Frisian-Dutch group was outperformed by the monolinguals on irregular inflection, no such effect was found for the Arabic-Dutch and Tarifit-Dutch groups. We suggest that both limited impact of cross-language similarity and much impact of amount of Dutch explain these patterns. Explaining this below, we focus on irregulars, because these showed the most pronounced effects of group, and revealed patterns that contradicted the hypothesized effect of cross-linguistic distance.

The irregular items in the standardized test we used overlap less between Frisian and Dutch than the regular items (Blom and Bosma 2016). As a consequence, Frisian-Dutch bilingual children do not benefit from Frisian when they use Dutch irregulars. For example, all regular noun plurals end in Dutch and Frisian in the same suffix, whereas all four irregular noun plurals differ between the two languages, which means that Frisian interferes (Blom and Bosma 2016). The second effect concerns the role of amount of Dutch. The model with only bilinguals showed that more use of Dutch at home predicted higher inflection accuracy. The impact of use of Dutch at home is in line with other research showing effects of input (see for an overview of input effects in bilingual children: Unsworth 2016), and supports a UB approach to inflection that foregrounds experiential effects. As shown in Table 1, in the Frisian-Dutch group, Dutch is used relatively infrequently, whereas in the Arabic-Dutch and Tarifit-Dutch groups, Dutch is used relatively frequently. Firstly, this pattern confirms findings in other research showing that migrant families from Moroccan descent use more Dutch at home than families from Turkish descent (Scheele et al. 2010). Secondly, it explains why the Frisian-Dutch children do not outperform the Arabic-Dutch and Tarifit-Dutch children, and why irregulars show, from the perspective of cross-linguistic

distance, an unexpected pattern. In particular, irregulars that are not highly frequent—which is the case for most items in this study—will require a substantial amount of input before these forms become entrenched and part of a child's repertoire, because they tend to rely on token frequency (Blom et al. 2012; Paradis et al. 2011; Rispens and De Bree 2015).

The effect of Group was moderated by regularity and parental education. For each of the groups, we found that higher parental education was associated with higher accuracy at using inflection, but for the Turkish-Dutch and Frisian-Dutch bilingual groups, the effect of parental education was stronger than for the Arabic-Dutch and Tarifit-Dutch groups (Figures A2 and A5). These differential patterns underscore that effects of parental education can vary across monolingual and bilingual populations and across different bilingual populations (Scheele et al. 2010). They also suggest that previous findings on the impact of socioeconomic status on Dutch vocabulary size in Turkish-Dutch children (Prevoo et al. 2014) hold for Dutch inflectional morphology as well. In a sample of Turkish-Dutch 5-year-old children, Prevoo et al. (2014) found that a higher socioeconomic status (which included parental education) predicted Dutch vocabulary. This relationship was explained by reading input (the availability of children's books at home).

We conclude that a binary categorical distinction between monolingual and bilingual groups is untenable. Factors that predict a higher accuracy at using (irregular) inflection in bilingual children comprise a smaller cross-linguistic distance, a larger amount of use of Dutch at home, and a higher level of parental education.

4.3. Verbal Short-Term (VSTM) and Verbal Working Memory (VWM)

We predicted that VSTM would be positively related to inflection accuracy, regardless of bilingual group and regularity, and that better VWM scores would predict higher accuracy at using regular but not irregular inflection. Overall, the results showed that better VSTM predicted better inflection accuracy, supporting results in previous research (Adams and Gathercole 1996, 2000; Baddeley et al. 1998; French and O'Brien 2008; Masoura and Gathercole 2005; Paradis 2011; Service 1992). However, a significant interaction effect between Group and VSTM showed that this effect held for all groups, except the Turkish-Dutch bilingual children. This interaction did not fit in model 2 with the bilinguals only, and prevented this model from converging. In our study, the Turkish-Dutch bilinguals scored, on average, lower on inflection than the other bilingual groups. This may suggest that the role of VSTM in learning inflection is dependent on level of development, similar to what has been proposed for word learning (Baddeley 2003). Interestingly, our findings differ from those of Verhagen and Leseman (2016), who found that VSTM predicted accuracy at using inflection in 5-year-old Turkish-Dutch children. The children in our sample were, on average, older. However, probably more relevant for understanding the differential findings is the fact that Verhagen and Leseman measured VSTM using (non-)word recall tests whereas we used a Forward Digit Span task. Based on several studies in which relationships between VSTM and vocabulary development were investigated, Baddeley (2003) concludes that digit span tasks tend to correlate less reliably with vocabulary than nonword recall tests. Model 2 did reveal a significant interaction between VSTM and parental education showing that the positive effect of better VSTM on inflection accuracy was found for lower, but not for higher levels of parental education. Parental education is a proxy for input quality and quantity (Hoff 2006), and the observed interaction indicates that children who experience higher input quality and quantity rely less on VSTM for learning inflection than children who receive less input and input of lower quality.

VWM did not predict children's accuracy at using inflection. A significant interaction was found between VWM and parental education, showing that for lower parental education, higher accuracies on inflection were associated with *lower* VWM. We do not have an explanation for this unexpected and counter-intuitive pattern, but it is relevant to note that with increasing parental educational level, the differences between higher and lower VWM outcomes almost disappeared or even reversed. The absence of an effect of VWM is not in agreement with the study of Verhagen and Leseman (2016), in which better VWM scores in

monolingual Dutch and bilingual Turkish-Dutch children predicted fewer errors at using inflection. One possible explanation for the contradicting results may have to do with the inclusion of non-verbal intelligence. Nonverbal intelligence, VSTM and VWM have all been found to predict bilingual children's accuracy at using inflection (Paradis 2011; Verhagen and Leseman 2016), but tend to be correlated (Colom et al. 2006; Engle et al. 1999; Engel de Abreu et al. 2010; Giofre et al. 2013). For example, in our study, VWM was moderately to strongly correlated with non-verbal intelligence and VSTM, respectively (Appendix B). While Paradis (2011) considered the simultaneous effects of VSTM and non-verbal intelligence, Verhagen and Leseman (2016) included the effects of VSTM and VWM. The results of our study indicate that when all three measures of children's language learning potential are included, only VSTM and non-verbal intelligence are relevant predictors.

In conclusion, children with better VSTM are more accurate at using inflection, but this effect does not generalize to all children. Analytical reasoning (indexed by non-verbal intelligence) may be more important than VWM for children's accurate use of inflection.

5. Limitations and Future Research

The study had a number of limitations. The different groups had unequal sizes and, particularly, the size of the Arabic-Dutch sample was very small. This group was not targeted and only during testing it turned out that the children were speakers of (Moroccan-)Arabic instead of Tarifit. Despite its small size, the behavioral patterns in this group of children lined up with patterns in the other groups, strengthening the conclusions. Another limitation is that for each construct, only one indicator was used. Multiple indicators enable the extraction of a latent variable, which provides a more robust reflection of the construct that is less dependent on one specific task. This would, for example, allow a more reliable distinction between VWM and non-verbal intelligence. In addition, the role of cognitive functions related to VWM, such as attention and inhibitory control, could be examined. In this study, we included proportional measures of Dutch input quantity at home. However, input quality is often more important than mere input quantity (Rowe 2012). Conceivably, measures of Dutch input quality could provide insight into the differential patterns of parental education across groups. In this study, VSTM was related to accuracy at using inflection, but this relationship was not found for all children (all except for Turkish-Dutch bilingual children, and bilingual children whose parents were more highly educated). A venue for future research concerns the questions which children do and do not benefit from good VSTM, and why this is the case.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Items

Table A1. Items of the TAK Word Formation Test (Verhoeven and Vermeer 2001). © 2001, by Cito, Arnhem, the Netherlands.

Noun plurals Regular		
bril-brillen	brɪl—'brɪlə	glasses-glasses
oor-oren	o:r—'o:rə	ear-ears
krant-kranten	krɑnt—'krɑntə	newspaper-newspapers
oog-ogen	o:x —'o:gə	eye-eyes
vlinder-vlinders	vlɪndər —'vlɪndərs	butterfly-butterflies
lepel-lepels	'le:pəl—'le:pəls	spoon-spoons
emmer-emmers	'ɛmər—'ɛmərs	bucket-buckets
trommel-trommels	'trɔməl—'trɔməls	drum-drums
Noun plurals Irregular		
weg-wegen	ʋɛx—'ʋe:gə	road-roads
dak-daken	dɑk—'da:kə	roof-roofs
slot-sloten	slɔt—'slo:tə	lock-locks
gat-gaten	ɣɑt—'ɣa:tə	hole-holes
Past participles Regular		
kook-gekookt	ko:k— ɣə'ko:kt	cook-cooked
plak-geplakt	plɑk— ɣə'plɑkt	paste-pasted
speel-gespeeld	spe:l—ɣə'spe:lt	play-played
fiets-gefietst	fi·ts—ɣə'fi·tst	bike-biked
Past participles Irregular		
zit-gezeten	zɪt—ɣə'ze:tə	sit-sat
vlieg-gevlogen	vli·x—ɣə'vlo:gə	fly-flown
kijk-gekeken	kɛik—ɣə'ke:kə	look-looked
drink-gedronken	drɪŋk—ɣə'drɔŋkə	drink-drunk
breng-gebracht	brɛŋ—ɣə'brɑxt	bring-brought
zoek-gezocht	zu·k—ɣə'zɔxt	seek-sought
verlies-verloren	vər'li·s—vər'lo:rə	lose-lost
koop-gekocht	ko:p—ɣə'kɔxt	buy-bought

Appendix B. Correlations between Fixed-Effect Variables

Table A2. Correlation coefficients between predictor variables. The lower-left to the diagonal is for all children; the higher-right to the diagonal is for the bilingual children only.

	Age	EdPar	Use Dutch	WNV	VSTM	VWM	Lemma Freq
Age	-	−0.047	-	0.019	0.279	0.328	0.001
EdPar	−0.038	-	-	0.239	0.212	0.168	0.000
Use Dutch	−0.089	−0.082	-	-	-	-	-
WNV	0.023	0.222	−0.141	-	0.232	0.381	0.000
VSTM	0.248	0.172	−0.108	0.204	-	0.610	0.001
VWM	0.321	0.152	−0.140	0.407	0.611	-	0.000
Lemma freq	0.001	0.000	0.000	0.000	0.001	0.000	-

Note: EdPar = parental education, WNV = non-verbal intelligence, VSTM = verbal short-term memory, VWM = verbal working memory, Lemma freq = lemma frequency.

The variables were averaged per subject and word (random factors) before calculating the correlation coefficients, which resulted in $n = 5232$ cases for all children and $n = 4128$ cases for the bilingual children only. In both groups, the highest correlation was found between VSTM and VWM.
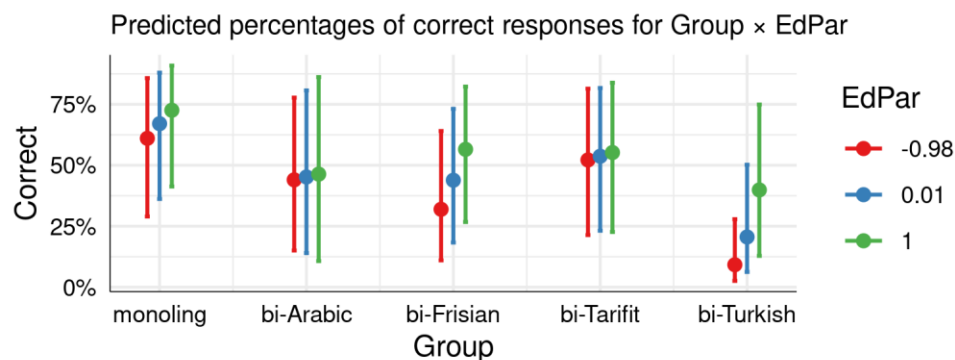
Appendix C. Interaction Graphs: All Children

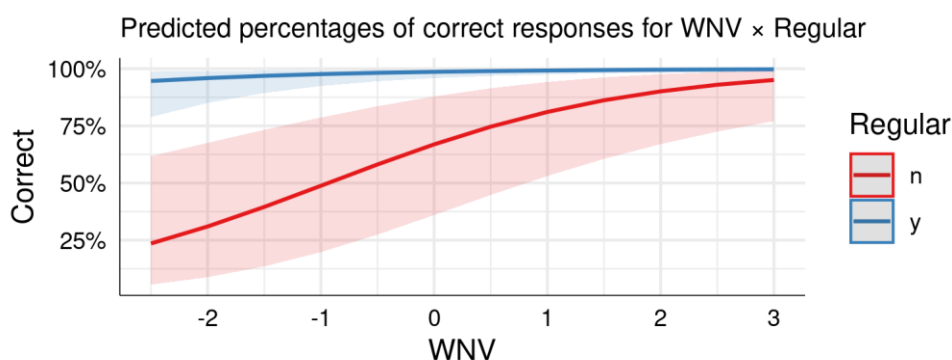**Figure A1.** Interaction between group and parental education (EdPar).

Figure A2. Interaction between non-verbal intelligence (WNV) and regularity of inflection.

Appendix D. Interaction Graphs: Bilingual Children
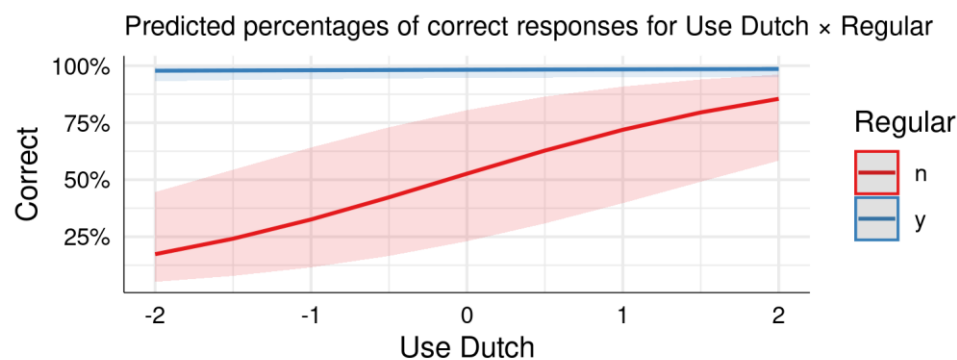
Figure A3. Interaction between use of Dutch at home and regularity of inflection.
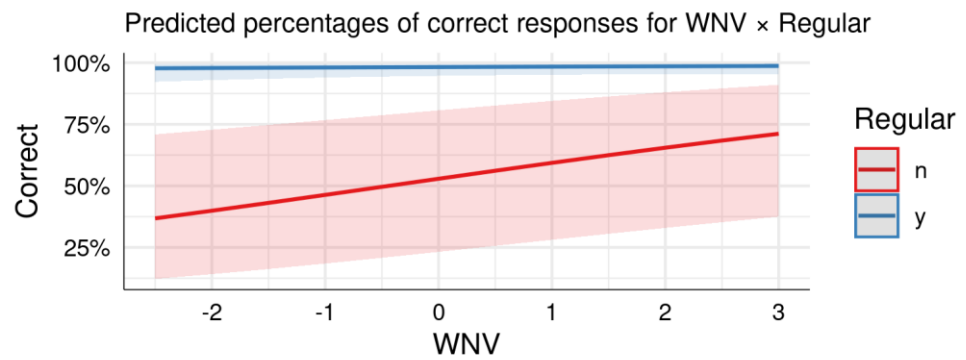
Figure A4. Interaction between non-verbal intelligence (WNV) and regularity of inflection.
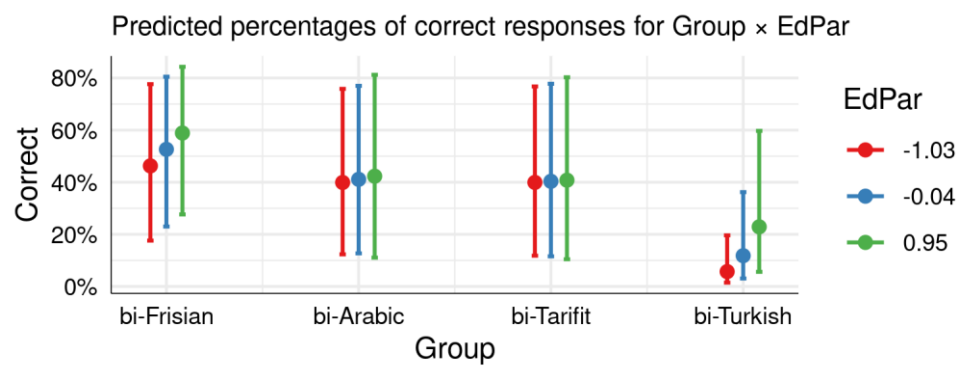

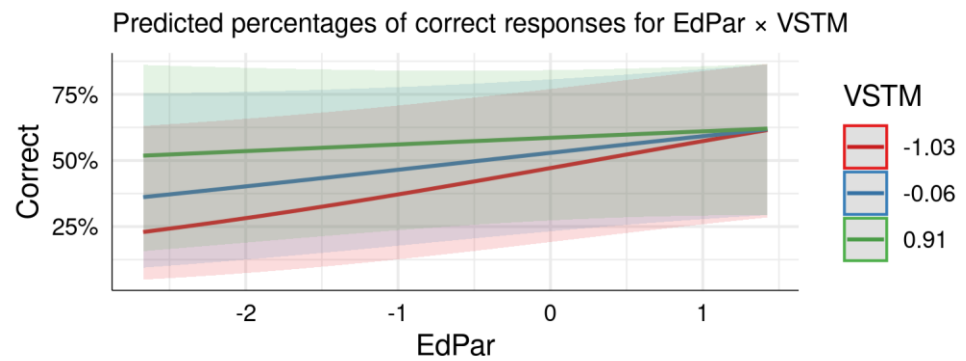**Figure A5.** Interaction between group and parental education (EdPar).

Figure A6. Interaction between parental education (EdPar) and verbal short-term memory (VSTM).
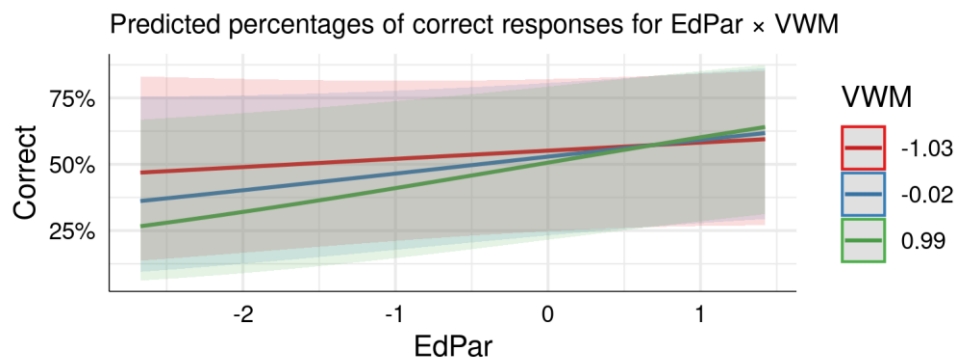

Figure A7. Interaction between parental education (EdPar) and verbal working memory (VWM).

References

Adams, Anne-Marie, and Susan E. Gathercole. 1996. Phonological working memory and spoken language development in young children. *Quarterly Journal of Experimental Psychology A* 49: 216–33. [CrossRef]

Adams, Anne-Marie, and Susan E. Gathercole. 2000. Limitations in working memory: Implications for language development. *International Journal of Language Communication Disorders* 35: 95–116. [PubMed]

Alloway, Tracy P. 2012. *Alloway Working Memory Assessment 2 (AWMA-2)*. London: Pearson.

Alloway, Tracy P., Susan E. Gathercole, Hannah Kirkwood, and Julian Elliott. 2008. Evaluating the validity of the automated working memory assessment. *Educational Psychology* 28: 725–34. [CrossRef]

Baddeley, Alan. 2003. Working memory and language: An overview. *Journal of Communication Disorders* 36: 189–208. [CrossRef]

Baddeley, Alan, Susan Gathercole, and Costanza Papagno. 1998. The phonological loop as a language learning device. *Psychological Review* 105: 158–73. [CrossRef]

Bates, Douglas, Martin Mächler, Benjamin M. Bolker, and Steven C. Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67: 1–48. [CrossRef]

Bialystok, Ellen, Fergus Craik, and Gigi Luk. 2012. Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences* 16: 240–50. [CrossRef]

Blom, Elma. 2019. Domain-general cognitive ability predicts bilingual children's receptive vocabulary in the majority language. *Language Learning* 69: 292–322. [CrossRef]

Blom, Elma, and Evelyn Bosma. 2016. The sooner the better? An investigation into the role of age of onset and its relation with transfer and exposure in bilingual Frisian-Dutch children. *Journal of Child Language* 43: 581–607. [CrossRef]

Blom, Elma, and Johanne Paradis. 2013. Past tense production by English second language learners with and without language impairment. *Journal of Speech, Language and Hearing Research* 56: 281–94. [CrossRef]

Blom, Elma, Johanne Paradis, and Tamara Sorenson Duncan, 2012. Effects of input properties, vocabulary size and L1 on the development of third person singular –s in child L2 English. *Language Learning* 62: 965–94. [CrossRef]

Blom, Elma, Tessel Boerma, Evelyn Bosma, Leonie Cornips, Kirsten van den Heuij, and Mona Timmermeister. 2020. Cross-language distance influences receptive vocabulary outcomes of bilingual children. *First Language* 40: 151–71. [CrossRef]

Boerma, Tessel, Frank Wijnen, Paul Leseman, and Elma Blom. 2017. Grammatical morphology in monolingual and bilingual children with and without language impairment: The case of Dutch plurals and past participles. *Journal of Speech, Language, and Hearing Research* 60: 2064–80. [CrossRef]

Booij, Geert. 1995. *The Phonology of Dutch*. Oxford: Clarendon Press.

Bosma, Evelyn, Wilbert Heeringa, Eric Hoekstra, Arjen Versloot, and Elma Blom. 2017. Verbal working memory is related to the acquisition of cross-linguistic phonological regularities. *Frontiers in Psychology* 8: 1487. [CrossRef]

Bretz, Frank, Torsten Hothorn, and Peter Westfall. 2016. *Multiple Comparisons Using R*. Boca Raton: CRC Press.

Bybee, Joan. 2007. *Frequency of Use and the Organization of Language*. Oxford: Oxford University Press.

Bybee, Joan. 2008. Usage-based grammar and second language acquisition. In *Handbook of Cognitive Linguistics and Second Language Acquisition*. Edited by Peter Robinson and Nick C. Ellis. New York: Routledge, pp. 216–36.

Cohen-Mimran, Ravit, Jasmeen Adwan-Mansour, and Shimon Sapir. 2013. The effect of morphological complexity on verbal working memory: Results from Arabic speaking children. *Journal of Psycholinguistic Research* 42: 239–53. [CrossRef]

Colom, Roberto, Irene Rebollo, Francisco. J. Abad, and Pei Chun Shih. 2006. Complex span tasks, simple span tasks, and cognitive abilities: A reanalysis of key studies. *Memory and Cognition* 34: 158–71. [CrossRef] [PubMed]

Den Ridder, Josje, Edith den Josten, Jeroen Boelhouwer, and Crétien van Campen. 2020. *De Sociale Staat van Nederland 2020*. The Hague: Sociaal en Cultureel Planbureau (SCP).

Dixon, L. Quentin, Shuang Wu, and Ahlam Daraghmeh. 2012. Profiles in bilingualism: Factors influencing kindergartners' language proficiency. *Early Childhood Education Journal* 40: 25–34. [CrossRef]

Engel de Abreu, Pascale M. J., and Susan E. Gathercole. 2012. Executive and phonological processes in second-language acquisition. *Journal of Educational Psychology* 104: 974–86. [CrossRef]

Engel de Abreu, Pascale M. J., Andrew R. A. Conway, and Susan E. Gathercole. 2010. Working memory and fluid intelligence in young children. *Intelligence* 38: 552–61. [CrossRef]

Engle, Randall W., Stephen W. Tuholski, James E. Laughlin, and Andrew. R. A. Conway. 1999. Working memory, short-term memory and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology General* 128: 309–31. [CrossRef]

Floccia, Caroline, Thomas Sambrook, Claire Delle Luche, Rosa Kwok, Jeremy Goslin, Laurence White, Allegra Cattani, Emily Sullivan, Kirsten Abbot-Smith, Andrea Krott, and et al. 2018. Vocabulary of 2-year-olds learning English and an additional language: Norms and effects of linguistic distance. *Monographs of the Society for Research in Child Development* 83: 30–42. [CrossRef] [PubMed]

Flom, Peter L., and David L. Cassell. 2007. Stopping stepwise: Why stepwise and similar selection methods are bad, and what you should use. Paper presented at the NorthEast SAS Users Group Inc 20th Annual Conference, Baltimore, MA, USA, November 11–14; vol. 11.

Fox, John, and Sanford Weisberg. 2019. *An R Companion to Applied Regression*, 3rd ed. Thousand Oaks: Sage.

Francot, Ryanne, Elma Blom, Martine Broekhuizen, and Paul Leseman. 2020. Profiles of bilingualism in early childhood: A person-centred Latent Profile Transition Approach. *Bilingualism: Language and Cognition*, 1–14. [CrossRef]

French, Leif M., and Irena O'Brien. 2008. Phonological memory and children's second language grammar learning. *Applied Psycholinguistics* 29: 463–87. [CrossRef]

Giofre, David, Irene C. Mammarella, and Cesare Cornoldi. 2013. The structure of working memory and how it relates to intelligence in children. *Intelligence* 41: 396–406. [CrossRef]

Hoff, Erika. 2006. How social contexts support and shape language development. *Developmental Review* 26: 55–88. [CrossRef]

Hoff, Erika. 2013. *Language Development*, 5th ed. Boston: Cengage Learning.

Hogan-Brun, Gabrielle, and Stefan Wolff. 2003. Minority languages in Europe: An introduction to the current debate. In *Minority Languages in Europe*. Edited by Gabrielle Hogan-Brun and Stefan Wolff. London: Palgrave Macmillan, pp. 3–15.

Kormos, Judit, and Anna Sáfár. 2008. Phonological short-term memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition* 11: 261–71. [CrossRef]

Kroll, Judith, and Ellen Bialystok. 2013. Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology* 25: 497–514. [CrossRef] [PubMed]

Lalleman, Josine A. 1986. *Dutch Language Proficiency of Turkish Children Born in the Netherlands*. Berlin and Boston: De Gruyter Mouton.

Lüdecke, Daniel. 2018. ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software* 3: 772.

Magimairaj, Beula M., and James W. Montgomery. 2012. Children's verbal working memory: Relative importance of storage, general processing speed, and domain-general controlled attention. *Acta Psychologica* 140: 196–207. [CrossRef]

Masoura, Elvira V., and Susan E. Gathercole. 2005. Phonological short-term memory skills and new word learning in young Greek children. *Memory* 13: 422–29.

McDonald, Janet L. 2000. Grammaticality judgments in a second language: Influences of age of acquisition and native language. *Applied Psycholinguistics* 21: 395–423. [CrossRef]

Mercator European Research Centre on Multilingualism and Language Learning. 2007. *The Frisian Language in Education in the Netherlands*. Leeuwarden: Fryske Akademy.

Nash, John, and Ravi Varadhan. 2011. Unifying optimization algorithms to aid software system users: Optimx for R. *Journal of Statistical Software* 43: 1–14. [CrossRef]

Nicoladis, Elena, and Johanne Paradis. 2012. Acquiring regular and irregular past tense morphemes in English and French: Evidence from bilingual children. *Language Learning* 62: 170–97. [CrossRef]

Nicoladis, Elena, Andrea Palmer, and Paula Marentette. 2007. The role of type and token frequency in using past tense morphemes correctly. *Developmental Science* 10: 237–54. [CrossRef]

Oller, D. Kimbrough, and Rebecca E. Eilers. 2002. *Language and Literacy in Bilingual Children*. Clevedon: Multilingual Matters.

Oostdijk, Nelleke. 2000. The Spoken Dutch Corpus Project. *The ELRA Newsletter* 5: 4–8.

Paradis, Johanne. 2011. Individual differences in child English second language acquisition. *Linguistic Approaches to Bilingualism* 1: 213–37. [CrossRef]

Paradis, Johanne, Elena Nicoladis, Martha Crago, and Fred Genesee. 2011. Bilingual children's acquisition of the past tense: A Usage-Based Approach. *Journal of Child Language* 38: 554–78. [CrossRef]

Paradis, Johanne, Yasemin Tulpar, and Antti Arppe. 2016. Chinese L1 children's English L2 verb morphology over time: Individual variation in long-term outcomes. *Journal of Child Language* 43: 553–80. [CrossRef]

Prevoo, Mariëlle, Maike Malda, Judi Mesman, Rosanneke Emmen, Nihal Yeniad, Marinus H. van IJzendoorn, and Mariëlle Linting. 2014. Predicting ethnic minority children's vocabulary from socioeconomic status, maternal language and home reading input: Different pathways for host and ethnic language. *Journal of Child Language* 41: 963–84. [CrossRef]

Rispens, Judith, and Elise De Bree. 2015. Bilingual children's production of regular and irregular past tense morphology. *Bilingualism: Language and Cognition* 18: 290–303. [CrossRef]

Rizaoğlu, Filiz, and Ayşe Gürel. 2020. Second language processing of English past tense morphology: The role of working memory. *International Review of Applied Linguistics in Language Teaching*. [CrossRef]

Rowe, Meredith L. 2012. A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development* 83: 1762–74. [CrossRef] [PubMed]

Scheele, Anna, Paul Leseman, and Aziza Mayo. 2010. The home language environment of mono- and bilingual children and their language proficiency. *Applied Psycholinguistics* 31: 117–40. [CrossRef]

Schwartz, Mila, Ely Kozminsky, and Mark Leikin. 2009. Delayed acquisition of irregular inflectional morphology in Hebrew in early sequential bilingualism. *International Journal of Bilingualism* 13: 501–22. [CrossRef]

Service, Elisabet. 1992. Phonology, working memory, and foreign language learning. *Quarterly Journal of Experimental Psychology A* 45: 21–50. [CrossRef]

Surrain, Sarah, and Gigi Luk. 2017. Describing bilinguals: A systematic review of labels and descriptions used in the literature between 2005–2015. *Bilingualism: Language and Cognition* 22: 1–15. [CrossRef]

Tuller, Laurice. 2015. Clinical use of parental questionnaires in multilingual contexts. In *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*. Edited by Sharon Armon-Lotem, Jan de Jong and Natalia Meir. Bristol: Multilingual Matters, pp. 301–30.

Unsworth, Sharon. 2016. Quantity and quality of language input in bilingual language development. In *Language and the Human Lifespan Series. Bilingualism across the Lifespan: Factors Moderating Language Proficiency*. Edited by Elena Nicoladis and Simona Montanari. Washington: American Psychological Association, pp. 103–21.

Verhagen, Josje, and Paul Leseman. 2016. How do verbal short-term memory and working memory relate to the acquisition of vocabulary and grammar? A comparison between first and second language learners. *Journal of Experimental Child Psychology* 141: 65–82. [CrossRef]

Verhoeven, Ludo, and Ant Vermeer. 1985. Ethnic group differences in children's oral proficiency in Dutch. In *Ethnic Minorities and Dutch as a Second Language*. Edited by Guus Extra and Ton Vallen. Dordrecht: Foris, pp. 105–31.

Verhoeven, Ludo, and Ant Vermeer. 2001. *Taaltest Alle Kinderen (TAK) [Language Assessment all Children]*. Arnhem: CITO.

Wechsler, David, and Jack A. Naglieri. 2006. *Wechsler Nonverbal Scale of Ability: WNV*. San Antonio: PsychCorp.