

ARTICLE

Cross-linguistic influence in simultaneous and early sequential bilingual children: a meta-analysis

Chantal VAN DIJK^{1*}, Elise VAN WONDEREN², Elly KOUTAMANIS¹,
Gerrit Jan KOOTSTRA¹, Ton DIJKSTRA³, and Sharon UNSWORTH¹

¹Radboud University, Centre for Language Studies, Netherlands, ²Radboud University, Netherlands, and ³Radboud University, Donders Institute for Brain, Cognition and Behaviour & Centre for Language Studies, Netherlands

*Corresponding author: Erasmusplein 1, room 8.16, 6525 HT Nijmegen, The Netherlands.
E-mail: c.n.vandijk@let.ru.nl

(Received 27 July 2020; revised 2 December 2020; accepted 2 April 2021)

Abstract

Although cross-linguistic influence at the level of morphosyntax is one of the most intensively studied topics in child bilingualism, the circumstances under which it occurs remain unclear. In this meta-analysis, we measured the effect size of cross-linguistic influence and systematically assessed its predictors in 750 simultaneous and early sequential bilingual children in 17 unique language combinations across 26 experimental studies. We found a significant small to moderate average effect size of cross-linguistic influence, indicating that cross-linguistic influence is part and parcel of bilingual development. Language dominance, operationalized as societal language, was a significant predictor of cross-linguistic influence, whereas surface overlap, language domain and age were not. Perhaps an even more important finding was that definitions and operationalisations of cross-linguistic influence and its predictors varied considerably between studies. This could explain the absence of a comprehensive theory in the field. To solve this issue, we argue for a more uniform method of studying cross-linguistic influence.

Keywords: cross-linguistic influence; bilingual children; surface overlap; morphosyntax; meta-analysis

Introduction

How a bilingual child's two languages affect each other has been a prominent topic of research in the field of bilingual first language acquisition over the past three decades. Such CROSS-LINGUISTIC INFLUENCE, most commonly investigated at the level of (morpho) syntax, has been attested in both the spontaneous and elicited speech production of simultaneous bilingual children, as well as in their comprehension and judgements of

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

sentences (see Serratrice, 2013, for an overview). Cross-linguistic influence is defined here as the overuse or overacceptance of (morpho)syntactic properties in bilingual children's one language under influence of their other language. For example, Italian-English bilingual children have been found to overuse overt subject pronouns in Italian and this has been argued to result from cross-linguistic influence from English (e.g., Serratrice, Sorace & Paoli, 2004). Researchers have aimed to identify the contexts in which cross-linguistic influence is most likely to appear. Well-studied predictors of cross-linguistic influence include surface overlap, language domain, language dominance, and age (e.g., Foroodi-Nejad & Paradis, 2009; Hulk & Müller, 2000; Müller & Hulk, 2001; Sorace & Serratrice, 2009; Yip & Matthews, 2000).

Evidence for the contribution of these predictors is mixed, however. Cross-linguistic influence is not always found when predicted (e.g., Argyri & Sorace, 2007; Nicoladis, 2002, 2003) and it is sometimes found when not predicted (e.g., Foroodi-Nejad & Paradis, 2009; Strik & Pérez-Leroux, 2011). Furthermore, cross-linguistic influence varies from child to child, as evidenced by the large standard deviations found in many studies (e.g., Mykhaylyk & Ytterstad, 2017; Nicoladis, 2006). As a consequence, there is neither consensus on the extent to which cross-linguistic influence in bilingual language acquisition takes place, nor what predicts it. To shed light on these issues, we conducted a meta-analysis to systematically examine the effect of morphosyntactic cross-linguistic influence in relation to surface overlap, language domain, language dominance, and age.

This paper is organised as follows. The next section discusses previous studies on cross-linguistic influence and the role of our predictors of interest. Then we list our research questions and hypotheses. The method section details our screening process, our coding procedure for surface overlap, language domain, language dominance and age, and how we calculated effect sizes for cross-linguistic influence. Subsequently, we present the outcomes of the meta-analysis and we discuss the results in relation to previous literature. Finally, we formulate recommendations for future studies based on our findings.

Morphosyntactic development in bilingual children

Research on cross-linguistic influence is embedded in a larger debate about the architecture of simultaneous bilingual children's language systems. In the pioneering work of the 1990s, researchers focussed on whether or not children's morphosyntactic systems developed independently from one another (e.g., de Houwer, 1990; Meisel, 1989; Paradis & Genesee, 1996). Taking separate systems as a starting point, research in the subsequent two decades investigated the extent to which cross-linguistic influence occurred (e.g., Hulk & Müller, 2000; Meisel, 2007; Paradis & Genesee, 1996; Serratrice, 2013).

Early work on cross-linguistic influence considered young children's spontaneous speech production in (multiple) case studies. Researchers typically compared the development of morphosyntactic properties in bilingual and monolingual children over a period of time (e.g., Döpke, 1998; Hulk & Müller, 2000; Paradis & Genesee, 1996). On the one hand, bilingual children were found to behave in language-specific ways, showing that they were able to differentiate the morphosyntactic rules of their languages (e.g., Döpke, 1998; Paradis & Genesee, 1996). On the other hand, the two languages were found to influence each other in both quantitative and qualitative ways: quantitative when acquisition of a certain morphosyntactic property was

facilitated or delayed in bilingual children under influence of their other language; and qualitative when bilingual children used a morphosyntactic property unattested in the speech of monolingual peers under influence of their other language (e.g., Müller & Hulk, 2001; Paradis & Genesee, 1996; Yip & Matthews, 2000).

More recent studies have typically employed experimental techniques, resulting in data on a wide range of linguistic properties and language combinations (see Serratrice, 2013 for an overview). These data have allowed researchers to systematically test for cross-linguistic influence under specific conditions in larger groups of bilingual children. Furthermore, they make it possible to study cross-linguistic influence not only on the basis of children's speech production, but also children's comprehension and judgements (e.g., Meroni, Smeets & Unsworth, 2017; Serratrice, 2007). At the same time, the comparison between bilingual and monolingual peers has remained central. Experimental studies have found similar patterns of behaviour as those using spontaneous speech data: bilingual children differentiated between the morphosyntactic properties of their languages, but at the same time showed quantitative and – to a lesser degree – qualitative cross-linguistic influence (e.g., Argyri & Sorace, 2007; Nicoladis, 2006; Strik & Pérez-Leroux, 2011).

Some studies have investigated cross-linguistic influence by comparing different groups of bilingual children with each other rather than comparing bilinguals with monolinguals (e.g., Kaltsa, Tsimpli & Argyri, 2019; Serratrice, Sorace, Filiaci & Baldo, 2009; Sorace & Serratrice, 2009). Such a design allows researchers to manipulate morphosyntactic properties cross-linguistically whilst at the same time controlling for bilingual vs. monolingual status (and all that this may entail) – we return to this design in more detail in the discussion. Because the vast majority of (experimental) studies on cross-linguistic influence have used a monolingual control group alongside a single bilingual group, we have focussed on this design in the present study.

Despite the many studies on the topic, the circumstances under which cross-linguistic influence emerges remain elusive. Cross-linguistic influence has been attested in various language combinations, for different linguistic properties, and using different tasks, but findings are inconsistent. Study outcomes can differ even when the same morphosyntactic property in the same language was under investigation (compare Rodina, Kupisch, Meir, Mitrofanova, Urek & Westergaard, 2020; Schwartz, Minkov, Dieser, Protassova, Moin & Polinsky, 2015). Various predictors of cross-linguistic influence have been identified to explain this variability. Typically, these have been discussed in relation to the PRESENCE of cross-linguistic influence – namely, whether certain conditions have to be met for cross-linguistic influence to occur – and in relation to the STRENGTH of cross-linguistic influence – namely, whether under certain circumstances the effect size of cross-linguistic influence increases.

In this study, we focus on four factors frequently studied in relation to cross-linguistic influence: (1) the type of surface overlap between bilingual children's languages, (2) the language domains involved, (3) language dominance, and (4) children's age. Whilst other factors, such as input quality (e.g., Paradis & Navarro, 2003) and economy principles (e.g., Gavarró, 2003; Serratrice et al., 2009), have also been argued to predict cross-linguistic influence, the number of studies investigating these variables is more limited and hence they are not included here. In the following four subsections, we discuss each of the factors of interest in more detail. We will end this section by discussing other reasons why there is such variation in results between and within studies on cross-linguistic influence.

Predictors of cross-linguistic influence

Surface overlap

One factor argued to predict the presence of cross-linguistic influence in bilingual children is the type of overlap between children's languages. According to Hulk and Müller (2000; Müller & Hulk, 2001), there has to be ambiguity in the child's language input for cross-linguistic influence to occur: if a certain structure in language A can be analysed (by the child) by either syntactic analysis X or Y and language B provides evidence for analysis X only, language B may reinforce the use of that analysis in language A, resulting in quantitative cross-linguistic influence. In other words, a certain type of overlap between children's languages is NECESSARY for cross-linguistic influence to occur (see Döpke, 1998 for a similar proposal). Hulk and Müller's overlap hypothesis is usually referred to in terms of surface or structural overlap. Whilst some authors make an explicit distinction between the two terms (e.g., Nicoladis, 2006; Schmitz, Patuto & Müller, 2012), most use them interchangeably to refer to the same construct. We use SURFACE OVERLAP throughout.

Hulk and Müller's overlap condition describes a situation of PARTIAL OVERLAP (e.g., Unsworth, 2003). There is optionality in language A – due to ambiguity in the input – and in language B one of these options is the preferred option. As a consequence, cross-linguistic influence is predicted to go unidirectionally from language B to language A. For example, in Persian, compounds can either be left- or right-headed (e.g., *bee-honey* for honeybee versus *headache*). In English, compounds can only be right-headed (e.g., Foroodi-Nejad & Paradis, 2009). As a consequence, in Persian–English bilingual children English may reinforce the use of right-headed compounds in Persian, leading to their overproduction. Following the surface overlap condition, however, situations of COMPLETE OVERLAP (i.e., where bilingual children's two languages behave identically) and NO OVERLAP (i.e., where they behave completely differently) should not result in cross-linguistic influence.

Whilst some scholars have found cross-linguistic influence in the direction predicted by surface overlap (e.g., Austin, 2007; Haznedar, 2007; Hulk & Müller, 2000), others have not; or they have found evidence of cross-linguistic influence in the absence of surface overlap (e.g., Argyri & Sorace, 2007; Foroodi-Nejad & Paradis, 2009; Nicoladis & Gavrilá, 2015).

Language domain

A second factor that has been argued to predict the presence of cross-linguistic influence in bilingual children is the language domain of the morphosyntactic property tested. Hulk and Müller proposed that, in addition to surface overlap, cross-linguistic influence only occurs in the domain where syntax interfaces with pragmatics, the so-called C-domain (e.g., Hulk & Müller, 2000; Müller & Hulk, 2001). An example is children's use of subject pronouns in a null subject language (e.g., Argyri & Sorace, 2007; Serratrice, 2007; Sorace, Serratrice, Filiaci & Baldo, 2009). Null subject languages allow both overt and null pronouns in subject position. However, the choice of a pronoun depends on discourse-pragmatics principles (e.g., Carminati, 2002). In particular, whilst a null pronoun is typically used to refer back to the topic of the discourse, an overt pronoun signals a shift in discourse topic. Consequently, subject pronoun use in null subject languages has been argued to be at the interplay of syntax and (discourse-)pragmatics (e.g., Sorace & Serratrice, 2009; Sorace *et al.*, 2009). However, Hulk and Müller did not rule out other domains at

the interface with syntax (e.g., Hulk & Müller, 2000; Müller & Hulk, 2001, p. 2). Non-interface areas, such as purely syntactic language properties, were predicted to be unaffected (e.g., compounding; Foroodi-Nejad & Paradis, 2009; Nicoladis, 2002; root infinitives; Hulk & Müller, 2000).

Whilst some researchers have found evidence for Hulk and Müller's proposal (e.g., Argyri & Sorace, 2007; Haznedar, 2007), others have also found cross-linguistic influence in other domains, especially syntax-semantics (e.g., genericity and specificity; Serratrice et al., 2009; indefinite object scrambling; Meroni et al., 2017). Furthermore, cross-linguistic influence has not always been attested when discourse pragmatics were involved (e.g., Argyri & Sorace, 2007). Moreover, cross-linguistic influence in purely (morpho)syntactic properties of language is also attested (e.g., Argyri & Sorace, 2007; Nicoladis, 2012; Strik & Pérez-Leroux, 2011).

Language dominance

A third factor that has been related to cross-linguistic influence is language dominance. Bilingual children typically have a dominant and a weaker language (e.g., Grosjean, 1982). What counts as a child's dominant language can be defined in various ways, for example, as the language a child is most proficient in (e.g., Unsworth, Chondrogianni & Skarabela, 2018). Language dominance has been observed to predict both the presence and the strength of cross-linguistic influence. Some studies have found cross-linguistic influence to be unidirectional and, thus, to predict the direction of cross-linguistic influence: namely, from children's dominant language into their non-dominant language (e.g., Argyri & Sorace, 2007; Yip & Matthews, 2000). Others have shown cross-linguistic influence to be bidirectional and to be present regardless of languages' dominance status. However, some studies found language dominance to predict the strength of cross-linguistic influence. To be more precise, the weaker the language was children have been tested in, the stronger the effect of cross-linguistic influence (e.g., Foroodi-Nejad & Paradis, 2009; Kidd, Chan & Chiu, 2015; Nicoladis, 2006). At the same time, others have found no effects of language dominance (e.g., Foroodi-Nejad & Paradis, 2009; Nicoladis, 2002; Unsworth, 2012).

Age

A final factor observed to affect the presence and strength of cross-linguistic influence is age. Earlier studies of cross-linguistic influence were typically corpus studies with very young bilingual children (often before the age of four) investigating the development of a certain morphosyntactic property over a longer period of time (e.g., Döpke, 1998; Müller & Hulk, 2001; Serratrice et al., 2004). As already discussed, in those studies cross-linguistic influence was evident during time periods where bilingual children's acquisition was slower or faster than monolingual peers', and where bilingual children used qualitatively different structures than monolingual peers. Importantly, these studies suggested that cross-linguistic influence is a developmental phenomenon that, with sufficient language exposure, disappears over time (e.g., Döpke, 1998; Müller & Hulk, 2001; Paradis & Genesee, 1996).

In more recent experimental work, researchers have explored cross-linguistic influence in older bilingual children (e.g., Daskalaki, Chondrogianni, Blom, Argyri & Paradis, 2019; Kaltsa et al., 2019). In an early study, Argyri and Sorace (2007) found evidence for cross-linguistic influence in seven-to-nine-year-old children, and others have found cross-linguistic influence to remain stable with age (e.g., Bosch & Unsworth, 2020; Nicoladis, 2002, 2003). This suggests that rather than being an

exclusively developmental phenomenon, cross-linguistic influence may be part and parcel of being bilingual (e.g., Nicoladis, 2006, 2012; Serratrice, 2013, 2016). At the same time, some experimental studies have found the effect of cross-linguistic influence to diminish (e.g., Serratrice *et al.*, 2009; Sorace *et al.*, 2009; Unsworth, 2012) or even increase with age (e.g., Nicoladis & Gavrilu, 2015). As a consequence, it is currently still unclear whether cross-linguistic influence is primarily a developmental phenomenon, mostly found in young bilingual children, or persists with age. Furthermore, as pointed out to us by an anonymous reviewer, age can be an index of language input and might therefore correlate with the (cumulative) amount of language exposure children receive in their two languages. We return to this latter point in the discussion.

In sum, despite or perhaps even because of the considerable body of experimental research on the topic, there is as yet no consensus about the circumstances under which cross-linguistic influence occurs. The presence of cross-linguistic influence and effects of its predictors vary across studies. In the next section, we discuss several explanations for this variability.

Accounting for variability across studies

First of all, study designs vary considerably in task set-up, morphosyntactic properties, and language pairs tested. Furthermore, the context of bilingual acquisition varies both within and across studies (e.g., in terms of input and age of onset). Whilst this variation across studies is necessary to detect whether there is a robust effect of cross-linguistic influence, study differences may influence the extent of cross-linguistic influence in unknown ways.

Second, surface overlap and language dominance have been defined and operationalized in many ways. With regard to surface overlap, some studies have based their predictions about surface overlap on the perspective of the adult language (e.g., Argyri & Sorace, 2007), whereas other studies focused on the (monolingual) child's point of view (e.g., Pirvulescu, Pérez-Leroux, Roberge, Strik & Thomas, 2014). For example, whilst adult native speakers of English might not allow left-headed compounds (e.g., *bee-honey* referring to the insect), monolingual children might consider such orders possible in English (e.g., Foroodi-Nejad & Paradis, 2009). The first scenario may have resulted in the underestimation of options available to the child and hence to the potentially incorrect classification of certain morphosyntactic properties as not overlapping between children's languages.

With regard to language dominance, authors have measured dominance differently, and operationalized it as both a categorical and continuous variable (e.g., Hervé, Serratrice & Corley, 2016; Nicoladis, 2002; Unsworth, 2012). For example, some divided bilingual children into dominance groups (e.g., Argyri & Sorace, 2007; Foroodi-Nejad & Paradis, 2009), whereas others included a continuous measure of dominance, such as percentage of language exposure or scores on some measure of language proficiency in their analyses (e.g., Bosch & Unsworth, 2020; Nicoladis, 2002). These differences in definitions and operationalizations may explain why studies have found different effects of surface overlap and language dominance.

Third, the absence of a significant effect in situations where cross-linguistic influence has been predicted should not be interpreted as absence of cross-linguistic influence. Instead, non-significant effects are to be expected due to random error. If we assume

that the power of studies investigating cross-linguistic influence is 80%, then there is a 20% chance that studies fail to detect a significant effect of cross-linguistic influence when it is in fact there. Scholars often interpret non-significant effects incorrectly as the absence of an effect (cf. Borenstein, Hedges, Higgins & Rothstein, 2009; Brysbaert, 2019). Instead, what is essential is whether the direction of the non-significant effects was consistent with cross-linguistic influence. Given that it is common for studies on cross-linguistic influence to test relatively few bilingual children (e.g., Foroodi-Nejad & Paradis, 2009; Strik & Pérez-Leroux, 2011), many studies probably even had a lower power level than 80%. Underpowered studies and random variables could therefore also explain why some studies have failed to find significant effects of surface overlap, language domain, language dominance and age, whilst others have.

The present study

The aim of the present study is to conduct a meta-analysis that systematically assesses cross-linguistic influence and its predictors. Such a meta-analysis allows us to go beyond problematic differences between studies, because summary effect sizes are calculated for relevant variables by averaging across studies. In this way, effects of cross-linguistic influence can be investigated in much larger groups of children than in individual studies. Furthermore, a meta-analysis can provide information on whether variation in the effect of cross-linguistic influence between studies appears to be random (i.e., is due to random error), or systematic (i.e., relates to predictor variables; Borenstein et al., 2009). Finally, a meta-analysis allows us to statistically test the role of predictor variables (e.g., Borenstein et al., 2009; Hedges & Olkin, 1985).

In this study, we address the following research questions:

RQ1 To what extent is there cross-linguistic influence in bilingual children at the level of morphosyntax and how consistent is this effect across studies?

Given that cross-linguistic influence has been attested in various studies (e.g., Serratrice, 2013), we expect to find an average effect size of cross-linguistic influence that is significantly larger than zero. At the same time, we expect considerable variation across studies due to differences in experimental designs. Nevertheless, findings from studies should generally be consistent with cross-linguistic influence.

RQ2 To what extent does surface overlap affect the strength and presence of cross-linguistic influence?

We hypothesize that if the strength of cross-linguistic influence is affected by surface overlap, its effect will be stronger in situations of partial surface overlap – when one language can reinforce a partially overlapping morphosyntactic structure in the other language – compared to situations without surface overlap. If, however, surface overlap is a NECESSARY condition for cross-linguistic influence to occur at all (e.g., Hulk & Müller, 2000), the effect of cross-linguistic influence will be significant only in situations with partial surface overlap and not in situations of no surface overlap.

RQ3 To what extent does language domain affect the strength and presence of cross-linguistic influence?

If language domain affects the strength of cross-linguistic influence, we expect cross-linguistic influence to be stronger for morphosyntactic properties that interact with discourse pragmatics compared to properties in other language domains. However, if the interaction between morphosyntax and discourse pragmatics is NECESSARY for cross-linguistic influence to be present (e.g., Hulk & Müller, 2000), the effect of cross-linguistic influence will only be significant in this domain and not in others.

RQ4 To what extent does language dominance affect the strength and presence of cross-linguistic influence?

If language dominance affects the strength of cross-linguistic influence (e.g., Argyri & Sorace, 2007; Foroodi-Nejad & Paradis, 2009), cross-linguistic influence should be stronger from children's dominant language into their non-dominant language than vice versa. If language dominance affects the presence of cross-linguistic influence, we hypothesize that cross-linguistic influence will be unidirectional from children's dominant language into the non-dominant language (e.g., Yip & Matthews, 2000). Hence, the effect of cross-linguistic influence should only be significant in children's non-dominant language and not in children's dominant language.

In sum, for the role of surface overlap (RQ 2), language domain (RQ 3) and language dominance (RQ 4), we formulated both a weaker and a stronger version of our hypotheses. The weaker hypothesis considers the predictor's effect on the strength of cross-linguistic influence. The stronger hypothesis considers its effect on the presence of cross-linguistic influence. We tested these hypotheses in two ways: (i) by using the authors' categorization of surface overlap, language domain and language dominance; and (ii) by categorizing the predictors ourselves. This second way of coding had the advantage, first of all, that it allowed for systematicity in terms of the definition and operationalization of cross-linguistic influence across studies; and, second, effect sizes could be taken into account for predictors not explicitly tested by the authors themselves.

RQ5 How does cross-linguistic influence develop with age?

We hypothesize that if cross-linguistic influence is a developmental phenomenon (e.g., Hulk & Müller, 2000; Paradis & Genesee, 1996), the effect of cross-linguistic influence should become weaker as children grow older. This is in line with studies that have found cross-linguistic influence to become weaker or disappear with age (e.g., Serratrice *et al.*, 2009; Sorace *et al.*, 2009). In contrast, if cross-linguistic influence is part and parcel of being bilingual, no significant effect of age on the strength of cross-linguistic influence should occur (e.g., Bosch & Unsworth, 2020; Nicoladis, 2002; 2003).

Method

Literature searches

We began by building a systematic inventory of studies investigating cross-linguistic influence in bilingual children (see Figure 1). We selected studies that measured differences in bilingual and monolingual children's performance on a certain language task for specific morphosyntactic properties and interpreted their findings in relation to cross-linguistic influence. The following additional inclusion and exclusion criteria were applied:

Inclusion criteria

- Children were simultaneous and early sequential bilinguals, i.e., age of onset for both languages was before the age of 4;0 (e.g., Genesee, Paradis & Crago, 2004; McLaughlin, 1978; Unsworth, 2013);
- Children were no older than 10;0 at the time of testing;
- The study presented original data.
- The study contained data from at least two bilingual and two monolingual children.

Exclusion criteria

- Studies with bimodal bilingual children, adoptees and children with a developmental language disorder;
- Priming and narrative studies.

We first searched Google Scholar for articles using various terms for cross-linguistic influence in combination with “bilingual children” (July, 2018; see Figure 1). We selected the first 980 returns for each term. In a second step, all articles were screened by two coders, on the basis of titles and abstracts with respect to aforementioned criteria. Subsequent full-text screening revealed that the vast majority of articles were irrelevant for our purposes because they either focussed on bilingual adults or on a topic other than cross-linguistic influence. In cases of disagreement, a third person acted as arbiter. If necessary, we contacted the study’s authors to check whether our criteria were met. In a third step, we searched the references cited in the selected articles for additional relevant studies, and we asked a number of experts in the field whether they knew of any studies not yet included.

In total, our search yielded 37 studies that met our inclusion criteria, and for 28 of these, we contacted authors for additional data (see below). In 15 cases, our request was met. For one study (Nicoladis, 2002), we were able to deduce the necessary information from reported statistics. For another study (Sorace et al., 2009), we estimated data from figures reported in the paper. For 11 studies, no sufficient data could be retrieved. Our final dataset therefore consisted of 26 studies.

Data coding

All but one of the 26 studies reported multiple comparisons between bilingual and monolingual children. For example, some studies investigated cross-linguistic influence in both bilingual children’s languages or for various morphosyntactic properties. Furthermore, some studies explored the behaviour of various bilingual groups, split up, for example, by age, country of residence, language dominance profile, and age of first exposure (e.g., Argyri & Sorace, 2007; Meir, Walters & Armon-Lotem, 2017; Serratrice et al., 2009; Serratrice, Sorace, Filiaci & Baldo, 2012; Strik & Pérez-Leroux, 2011). We entered each comparison as a separate row in a spreadsheet, yielding 187 unique datapoints.¹

¹In some situations not all comparisons reported in the selected studies met our initial selection criteria, either because a bilingual group was added as control group for another bilingual group, rather than as a test case of cross-linguistic influence (i.e., the Spanish-Dutch bilingual group in Serratrice et al., 2009; 2012; Sorace et al., 2009) or because a specific condition was not at the level of morphosyntax (i.e., the stressed

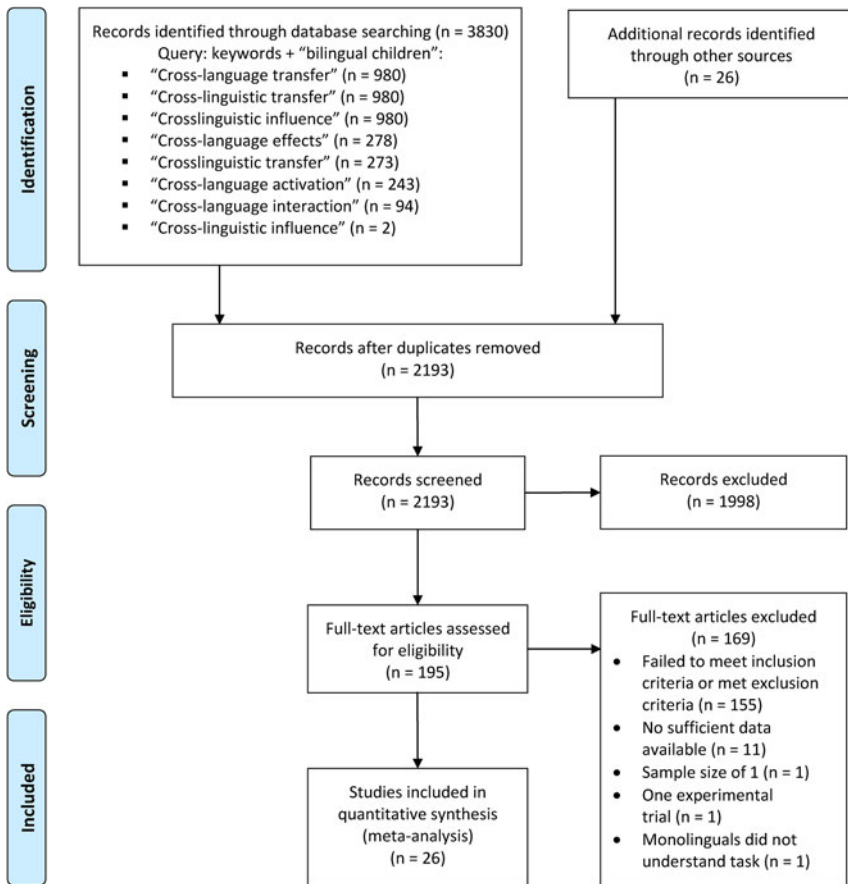


Figure 1. Flow chart showing selection process of experimental studies on cross-linguistic influence in simultaneous bilingual children following Preferred Reporting Items for Systematic Reviews and Meta-Analyses by Moher, Liberati, Tetzlaff, Altman & Group, 2009.

Subsequently, we coded each datapoint for a number of characteristics, including task design, language tested and morphosyntactic property, adapting a template provided by Metalab (<http://metalab.stanford.edu>; e.g., Bergmann, Tsuji, Piccinini, Lewis, Braginsky, Frank & Cristia, 2018). We coded for our variables of interest – namely, surface overlap, language domain, language dominance and age – and indicated whether a datapoint was considered as a testcase of cross-linguistic influence. The complete dataset is publicly available in the Data Archiving and Networked Services (DANS) repository (van Dijk, van Wonderen, Koutamanis, Kootstra & Unsworth, 2021).

and unstressed pronouns in English in Serratrice et al., 2012). Datapoints belonging to such comparisons were excluded.

Testcases of cross-linguistic influence

A datapoint was coded as a testcase of cross-linguistic influence in two steps. We coded first whether authors made explicit predictions about cross-linguistic influence (yielding 145 datapoints), and second, the direction of the predicted effect. For example, Foroodi-Nejad and Paradis (2009) elicited the production of compounds in Persian–English bilingual children. They predicted that if cross-linguistic influence were to take place, the children should use more right-headed compounds in Persian and/or more left-headed compounds in English compared to their monolingual peers. Hence, for the Persian task the direction of cross-linguistic influence predicted by the authors was coded as “more right-headed compounds” and for the English task as “more left-headed compounds”. Cross-linguistic influence was predicted for a total of 103 datapoints. For 42 datapoints, authors predicted no effect of cross-linguistic influence. This was typically the case when bilingual children’s languages patterned similarly for the morphosyntactic property under study (i.e., complete overlap). Hence, in those situations, bilingual children were predicted to behave similarly to monolingual children and datapoints were not included in the analyses.² Unfortunately, authors did not always formulate explicit predictions about cross-linguistic influence for each possible comparison (42 datapoints; e.g., Gathercole, Laporte & Thomas, 2005; Sorace et al., 2009).

To avoid inconsistencies across studies, we therefore applied a second, more neutral way of coding for testcases of cross-linguistic influence. We first identified every datapoint for which the authors had made no explicit predictions about cross-linguistic influence or for which they predicted no cross-linguistic influence (84 datapoints). We then checked whether the morphosyntactic property involved differed between bilingual children’s languages. This was done based on information that was provided in the articles. If a morphosyntactic property was identical between bilingual children’s languages, we predicted no cross-linguistic influence. These datapoints were then excluded from the analyses.³ If a morphosyntactic property differed between bilingual children’s languages, we coded the datapoint as a testcase of cross-linguistic influence. With regard to the direction of cross-linguistic

²Sometimes authors stated multiple – conflicting – hypotheses for the same datapoint (23 datapoints). For example, Serratrice and colleagues (2009) predicted unidirectional cross-linguistic influence from Italian to English based on Hulk and Müller’s surface overlap condition (e.g., Hulk & Müller, 2000; Müller & Hulk, 2001). However, they also formulated an alternative hypothesis based on economy considerations, which predicted unidirectional cross-linguistic influence in the opposite direction: from English to Italian. In addition, some authors predicted cross-linguistic influence according to one theory and no cross-linguistic influence according to another theory (e.g., Serratrice et al., 2009). In all conflicting situations, we categorized datapoints as a testcase of cross-linguistic influence (in the direction(s) indicated by the authors).

³Initially, we wanted to compare the average effect size of cases of complete overlap to testcases of cross-linguistic influence in order to shed light on the distinction between cross-linguistic influence and a more general effect of bilingualism. This turned out to be impossible, however, because the direction of individual effect sizes differs: for testcases of cross-linguistic influence the direction of Hedges’ *g* can be positive (consistent with cross-linguistic influence) or negative (inconsistent with cross-linguistic influence), whereas for cases of complete overlap there is no such distinction. Consequently, effect sizes would either always be positive or negative for cases of complete overlap. As a result, we deemed a comparison between cases of complete overlap and testcases of cross-linguistic influence to be uninformative. For a similar reason, situations with *complete overlap* were not included in the surface overlap analyses. Even if bilingual children would be found to behave differently from their monolingual peers in *complete overlap* situations, the effect size will never be positive (indicating cross-linguistic influence).

influence for these newly identified datapoints, we predicted that bilingual children would use a certain morphosyntactic structure more than their monolingual peers if this structure was preferred in their other language. Our second way of coding yielded 40 possible testcases of cross-linguistic influence in addition to those datapoints for which the authors themselves predicted cross-linguistic influence. We now turn to how we coded our moderator variables.

Predictors of cross-linguistic influence

Surface overlap

Our first predictor of interest was operationalized in two ways: (i) the authors' definition of overlap when based on Hulk and Müller's (2000) overlap hypothesis; and (ii) our own definition of overlap. The first operationalization yielded 35 datapoints that were identified by the authors as a situation of surface overlap, 60 situations of no surface overlap, six situations of complete surface overlap, and one situation where the authors first identified the situation as surface overlap but later argued that their task may in fact have tested a situation of complete overlap instead. From these 102 datapoints, we excluded datapoints for which no predictions could be made about cross-linguistic influence, i.e., when the predicted direction of cross-linguistic influence could not be inferred (6 datapoints) and when there was complete overlap between languages (6 datapoints). This left us with a total of 90 datapoints.

This way of coding turned out to have two disadvantages, however. First, many authors either did not explicitly discuss their study in relation to Hulk and Müller's overlap hypothesis (65 datapoints) or made no explicit predictions (20 datapoints). Second, for those datapoints that could be included in the analysis, authors varied as to whether they defined surface overlap in terms of (i) the adult- or the child-language system (we will elaborate on this in the Discussion); and (ii) a narrowly defined morphosyntactic context versus a broader context (see S1 for an explanation of narrow versus broad scope, Supplementary Materials). To deal with these issues, we recoded all datapoints using the same criteria: namely, based on the adult system and using a narrow scope. This not only allowed us to code for surface overlap in a uniform way, it also meant we could include datapoints from studies that made no explicit predictions about surface overlap. Datapoints were either coded as a situation of partial overlap (41 datapoints), a situation of no overlap (67 datapoints) or a situation of complete overlap (27 datapoints). For the remaining 52 datapoints, no unambiguous classification was possible. A detailed illustration of how datapoints were classified is provided in the Supplementary Materials (Supplementary Materials, S2).

Language domain

With respect to our second predictor of interest, language domain, we coded whether authors indicated which language domains were involved in the distribution of a certain morphosyntactic property: for example, syntax and pragmatics, or syntax and semantics. This was mentioned explicitly for 70 datapoints only: 43 datapoints involved discourse pragmatics, 20 datapoints were identified as purely (morpho) syntactic, and 7 datapoints involved semantics and not discourse pragmatics.

In an attempt to include more datapoints, we tried to systematically code for language domains ourselves. This turned out to be problematic. Hulk and Müller's (2000, p. 228) original definition was "the interface between two modules of grammar, and more particularly at the interface between pragmatics and syntax".

This definition is rather vague. Sorace and Serratrice (2009, p. 196) provide a more specific definition: “the distribution of the morphosyntactic construction of interest must be regulated by the interface with discourse pragmatics”. This latter definition can be straightforwardly applied to cases such as the distribution of null and overt subjects in languages such as Greek and Italian (e.g., Argyri & Sorace, 2007; Sorace et al., 2009), but in many other cases it was almost impossible to determine when discourse pragmatics were NOT involved. Hence, we decided to only analyse those datapoints for which language domain was mentioned by the authors.

Language dominance

Our third variable of interest, language dominance, was coded in two ways: (i) depending on the definition of the authors; and (ii) depending on the societal language of the bilingual children tested. The first way of coding was as follows: if authors classified a group of bilingual children as dominant in one of their two languages, we classified them as dominant in the target language (“target language”; 24 datapoints) or in the non-target language (“other language”; 23 datapoints) depending on which language was tested. Children considered balanced by the authors were coded as “balanced” (14 datapoints), and in cases where authors wrote that bilingual children’s dominance patterns varied, we coded dominance as “mixed” (6 datapoints). Information about dominance was not always provided. Consequently, language dominance was coded for a subset of datapoints only (67 in total).

Because language dominance was not consistently operationalized across studies, we also assessed children’s language profile in a more systematic way by coding whether or not the target language was also the language of the society where the bilingual children lived. Although we realize that this is only a rough proxy of language dominance (Hervé et al., 2016; Unsworth et al., 2018), it does provide a more objective measure of children’s language experience that could be coded for most studies.

Societal language was operationalized as the majority language of the country or area where the children were living and was coded as follows: if the societal language was the target language of a study, language dominance was coded as “target” (81 datapoints); if not, it was coded as “other” (97 datapoints). In one study (Hervé & Lawyer, unpublished manuscript) bilingual children came from different countries with different societal languages (“mixed”; 8 datapoints) and in one study (Nicoladis & Gavrilu, 2015) there was no clear distinction in status for the children’s two languages (“both”, 1 datapoint).

Age

Our fourth predictor of interest, age, was coded as mean age in months. In all studies the bilingual and monolingual children had similar mean ages except for the older bilingual group tested by Strik and Pérez-Leroux (2011). The age range of this bilingual group (6;05–7;11) and its monolingual control group (4;07–5;08) did not overlap. Such a large difference could have been problematic for our moderator analysis because younger children are typically less accurate on a language task than older children. Therefore, effects of cross-linguistic influence may both be exaggerated or minimized, depending on whether cross-linguistic influence is predicted to result in facilitation or delay. To avoid these effects, we excluded the datapoints from the older group of bilingual children in Strik and Pérez-Leroux (2011) from our analysis of age (4 datapoints). In addition, we also excluded the

results from the English task in Serratrice *et al.* (2009; 8 datapoints), because it was unclear which results belonged to the younger and older age group tested.

Effect sizes

Effect size estimates

We calculated the standardized effect size Hedges' g , and its variance, for each datapoint (e.g., Borenstein *et al.*, 2009; Hedges, 1981; all calculations were taken from Lakens, 2013, version 4.2). Each effect size was based on the differential mean of a bilingual and a monolingual group on a certain measure. The larger the difference in means between groups and the smaller their standard deviations were, the larger Hedges' g . In addition, we calculated the variance of Hedges' g . This indicated the precision of an effect size (e.g., Borenstein *et al.*, 2009). The larger the group sample sizes were, the smaller the variance and the more precise the corresponding effect size. In the meta-analysis, the more precise an effect size was, the more weight it was assigned.

The sign of the effect sizes indicated whether differences in scores found between bilingual and monolingual children were consistent with cross-linguistic influence. If the difference between a bilingual and a monolingual group was in the predicted direction the corresponding effect size was positive. If, on the other hand, there was a difference between a bilingual and monolingual group, but in the opposite direction than predicted (i.e., inconsistent with cross-linguistic influence), the corresponding effect size was negative. If bilingual and monolingual children had a similar score, the effect size was zero. We illustrate the interpretation of positive and negative effect sizes with two examples from Nicoladis (2006).

Nicoladis (2006) investigated cross-linguistic influence in adjective-noun orders in French–English bilingual children. In French, most adjectives typically appear postnominally (e.g., *une pomme vert*, “an apple green”) whereas some typically appear prenominally (e.g., *une grande pomme*, “a big apple”). In English, adjectives should – with a few exceptions – appear in prenominal position only (e.g., *a green/big apple*). Hence, Nicoladis predicted cross-linguistic influence from English into French to result in more prenominal adjectives in bilingual children's speech production compared to monolingual French peers. She elicited adjective-noun pairs in two conditions: (i) with typical French postnominal adjectives; and (ii) with typical French prenominal adjectives. She found bilingual children to produce the prenominal adjective order with postnominal adjectives in French more often than monolingual children. This difference between groups was consistent with cross-linguistic influence from English and therefore received a positive effect size (see Figure 1 in S4 in the Supplementary Materials, [Supplementary Materials](#)). In addition, bilingual children also placed prenominal adjectives more often in postnominal position than French monolingual children. This observation was inconsistent with cross-linguistic influence from English. Consequently, the effect size received a negative sign.

Data dependency

Effect sizes in our dataset were often not independent because they belonged to similar studies, similar groups of children or similar morphosyntactic properties investigated. Following Fernández-Castilla *et al.* (Fernández-Castilla, Jamshidi, Declercq, Beretvas, Onghena & Van den Noortgate, 2020; Fernández-Castilla, Maes, Declercq, Jamshidi, Beretvas, Onghena & Van den Noortgate, 2019) we controlled for dependencies in

the data by a multiple level cross-classified random effects model. In this model, we added three random effects for observation (i.e., an individual datapoint): namely, (i) a random intercept of observation nested in experimental task, which, in turn, was nested in data collection,⁴ (ii) a random intercept of observation nested in group of bilingual children, which, in turn, was nested in data collection, and (iii) a random intercept of observation nested in morphosyntactic property. All models in the paper used this random-effects structure.

Our random effect structure accounted for most dependencies in our dataset. One exception concerned those datapoints for which outcomes of different groups of bilingual children were compared to the same outcome from a group of monolingual children. To simplify our dataset, we collapsed means and standard deviations for datapoints belonging to different groups of bilingual children and similar groups of monolingual children by calculating weighted means and pooled standard deviations (e.g., Hoyt & Del Re, 2018). This resulted in a total of 128 datapoints. In our analyses of language dominance, we used uncollapsed datapoints in those situations where separate bilingual groups had different dominance patterns. This yielded 176 datapoints.

Data analyses

All analyses were conducted using the *rma.mv* function from the *metafor*-package (version 2.4-0; Viechtbauer, 2010) in R (version 3.6.3; R Core Team, 2020). For all analyses, the aforementioned random effect structure was applied. We performed two types of analyses: general analyses of the weighted mean effect size, and predictor analyses. First, we tested the average weighted mean effect size of cross-linguistic influence twice: (i) for those datapoints for which the authors made explicit predictions about the direction of an effect of cross-linguistic influence, and (ii) for all datapoints which we identified as possible testcases of cross-linguistic influence.

Second, we conducted separate moderator analyses with surface overlap, language domain and language dominance as predictors to investigate their effect on the strength and presence of cross-linguistic influence. With respect to surface overlap, effect sizes were compared twice: (i) between situations of surface overlap and no surface overlap as defined by the authors of the studies based on Hulk and Müller's (2000) overlap hypothesis, and (ii) between situations of partial overlap and no overlap as defined by us (see footnote 3 for an explanation why we could not take into account situations of complete overlap). If the difference between either of these surface overlap situations was significant, we tested whether surface overlap affected the presence of cross-linguistic influence. This was done by assessing whether the effect of no overlap and partial overlap was significantly larger than zero.

With respect to language domain, we conducted one analysis in which we compared the effect size of cross-linguistic influence for morphosyntactic properties that interacted with discourse pragmatics to the effect size for morphosyntactic properties that did not interact with discourse pragmatics. If this difference was significant, we assessed whether the effect size of cross-linguistic influence in each situation was significantly larger than zero.

⁴The same task in two different languages within the same data collection was coded as two separate tasks. In addition, we decided to nest participant groups and tasks in data collection rather than in study because data from Serratrice et al. (2009; 2012) and Sorace et al. (2009) were collected within the same data collection.

To test the effect of language dominance on the strength of cross-linguistic influence, effect sizes were compared twice: (i) between groups of children that were categorized as either dominant in the language tested or in the other language by the authors of the studies⁵; and (ii) between groups of children whose language of testing was the societal language and whose language of testing was not the societal language. If the difference between dominance categories was significant, we tested whether language dominance affected the presence of cross-linguistic influence. This was done by assessing whether the effect in the dominant and non-dominant language was significantly larger than zero.

Finally, with regard to age, a meta-regression was conducted with the mean age of the bilingual groups as continuous predictor of the effect size of cross-linguistic influence.

Results

Descriptive results

Our dataset consisted of 187 datapoints belonging to 750 unique bilingual children compared to 739 unique monolingual children. An overview of the characteristics of the studies in the dataset can be found in the Supplementary Materials (Supplementary Materials, S3). The majority of studies employed elicited production tasks. However, most observations in the dataset belonged to grammaticality judgement experiments. Only a few studies considered cross-linguistic influence in children's comprehension. There is considerable variation in the languages and linguistic properties tested. Although English has received most attention, there are many observations for other languages, too. Moreover, the language combinations under study were even more varied, with 17 unique language combinations. With regard to the linguistic properties tested, a large proportion investigated cross-linguistic influence in word order. Furthermore, quite a few studies focussed on null subjects and objects. However, the category with the most observations was genericity/specificity of plural noun phrases, even though only two studies tested for this property. Finally, with regard to the number of items tested per child, the majority of studies tested for cross-linguistic influence for a specific condition in less than 10 items. Eleven studies tested 6 items or less. Only six studies tested more than 20 items.

An overview of the characteristics of the bilingual groups in the dataset can be found in the Supplementary Materials (Supplementary Materials, S3) as well. The most frequently tested age group for bilingual children was on average four years old. Only two studies considered cross-linguistic influence in three-year-olds. With regard to the number of children studied, it is noteworthy that 17 studies compared groups of bilingual children to monolingual peers with a sample size of less than 20 for the bilinguals, and, in seven studies, with a sample size of less than 10. Although the majority of studies tested groups of 20 or more bilingual children, the majority of observations in our dataset belong to smaller sample sizes.

Cross-linguistic influence: average effect size and consistency

Figure 2 shows the datapoints per study for which either the authors or we predicted cross-linguistic influence (see S4 in the Supplementary Materials for forest plots with information about the morphosyntactic property and the language combination

⁵Datapoints belonging to children whose dominance profile was described as mixed or balanced were not included, due to low numbers of datapoints (mixed: 6; balanced: 14).

tested split out by task type, Supplementary Materials).⁶ The majority of effect sizes were larger than zero (73 datapoints), consistent with cross-linguistic influence. However, there was also a number of negative effect sizes (24 datapoints), which was inconsistent with cross-linguistic influence. Furthermore, the effect size of cross-linguistic influence varied between and within studies.

In our first analysis of the average effect size of cross-linguistic influence, we included only the 79 datapoints for which the authors of the studies explicitly predicted cross-linguistic influence. Effect sizes ranged from -1.24 to 2.66 . The random effects model revealed a significant small to medium average effect size of $g = 0.46$ ($(0.22, 0.71)$, $p < .001$). All models and model output reported in the result section can be found in the Supplementary Materials (Supplementary Materials, S5).

In our second analysis, we included an additional 34 effect sizes (a total of 113) previously identified as possible testcases of cross-linguistic influence. Now, the effect sizes ranged from -1.37 to 2.66 . The random effects model revealed a significant small to medium average effect size of $g = 0.39$ ($(0.21, 0.56)$, $p < .001$), slightly smaller than the average effect size in the first analysis.⁷

We further investigated the distribution of effect sizes in the second analysis using a funnel plot (Figure 3). In this plot, datapoints are plotted with their effect size on the x-axis and their standard error on the y-axis. The vertical line represents the average effect size. Datapoints with a smaller standard error are predicted to be scattered closer to the average effect size than datapoints with a greater standard error, as indicated by the diagonal lines. If studies with significant results are more likely to be published than studies with null results (publication bias), this should be reflected in an asymmetrical distribution of datapoints in the funnel plot: there should be more datapoints at the bottom right side of the distribution than at the bottom left side (e.g., Rothstein, Sutton & Borenstein, 2005). We do not see this distribution in Figure 3. Instead, there seemed to be some asymmetry in the opposite direction: namely, there were a number of effect sizes at the lower left side of the distribution.

Figure 3 also revealed quite some horizontal scatter of datapoints, a signal of heterogeneity in the data (e.g., Sterne, Sutton, Ioannidis, Terrin, Jones, Lau, Carpenter, Rücker, Harbord, Schmid, Tetzlaff, Deeks, Peters, Macaskill, Schwarzer, Duval, Altman, Moher & Higgins, 2011). This was confirmed by the significant test of heterogeneity of the model ($Q(112) = 505.00$, $p < .001$), which indicated that part of the variance in the data could not be explained by random error alone. This means that there must be other factors at play that account for differences in effect sizes. We tested whether this variance could be explained by our predictors of interest.

Analyses of predictors of cross-linguistic influence

We analysed the effect of our predictors by means of meta-regressions (e.g., Viechtbauer, 2010). All predictor analyses were conducted with positive effect sizes

⁶The distribution of the subset of effect sizes for which the authors explicitly predicted cross-linguistic influence was very similar to the distribution of effect sizes in Figure 2. Therefore, we decided to present the full set only.

⁷An anonymous reviewer was concerned that the average weighted effect size was not entirely reliable because we collapsed effect sizes of different task types. We did test for the effect for task type (elicited production, judgements and comprehension) in a moderator analysis, but this did not yield a significant effect. Outcomes of subset analyses for each task type can be found in the Supplementary Materials (Supplementary Materials, S4).

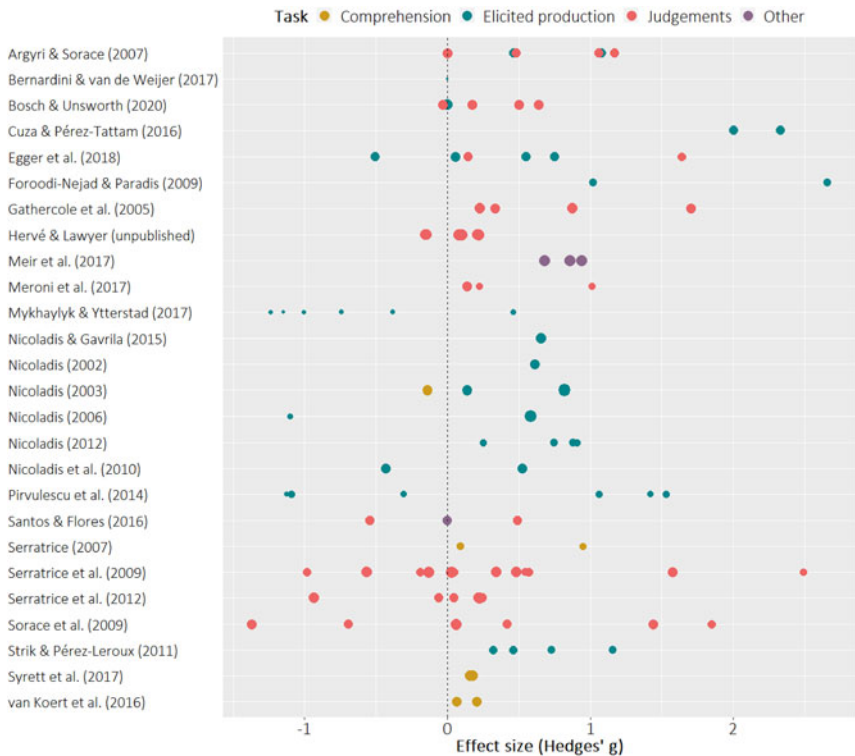


Figure 2. Effect sizes per study belonging to datapoints that were identified as possible testcases of cross-linguistic influence. The size of each dot indicates the precision of each effect size. Positive effect sizes reflect differences between bilingual and monolingual groups consistent with cross-linguistic influence. Negative effect sizes reflect differences inconsistent with cross-linguistic influence.

only. Negative effect sizes reflected divergent behaviour between bilingual and monolingual children that was inconsistent with cross-linguistic influence. We will discuss possible reasons for negative effects sizes in the Discussion. Regardless of what causes negative effect sizes in our dataset, interpreting them is difficult, and their presence might muddy our predictor analyses. Therefore, we decided to leave out negative effect sizes from further analyses. Moderator tests were conducted separately for our predictors of interest.

Surface overlap

The first analysis took into account those datapoints for which the authors made predictions about the presence or absence of cross-linguistic influence based on Hulk and Müller’s (2000) overlap hypothesis. Overall, the average effect size for surface overlap situations was slightly larger ($M = 0.69$, $SD = 0.81$, $range = 0-2.66$, $n = 20$) than the average effect size of situations without surface overlap ($M = 0.54$, $SD = 0.58$, $range = 0-2.49$, $n = 31$). However, this difference was not significant as shown by the moderator test of surface overlap ($Q_M(1) = 1.78$, $p = .182$).

The second analysis compared the average effect size of those datapoints that we identified as partial overlap situations versus no overlap situations. The average effect

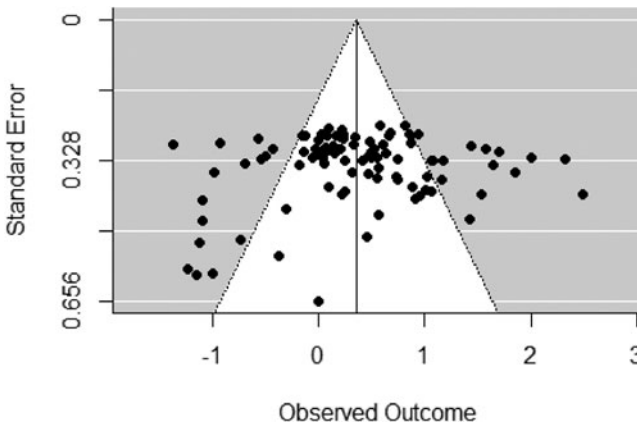


Figure 3. Funnel plot with observed effect sizes plotted on the x-axis and their standard errors on the y-axis.

size for partial overlap was slightly larger ($M = 0.76$, $SD = 0.72$, $range = 0-2.66$, $n = 17$) than the average effect size of no overlap ($M = 0.62$, $SD = 0.64$, $range = 0-2.49$, $n = 42$). However, the difference in effect size between partial overlap and no overlap situations did not reach significance either ($Q_M(1) = 0.37$, $p = .541$).

Language domain

The average effect size of morphosyntactic properties at the domain of discourse pragmatics ($M = 0.30$, $SD = 0.38$, $range = 0-1.17$, $n = 19$) was slightly smaller than the average effect size of morphosyntactic properties at other domains ($M = 0.39$, $SD = 0.40$, $range = 0-1.06$, $n = 17$). This difference was not significant, however ($Q_M(1) = 0.05$, $p = .832$).

Language dominance

In the first analysis, we compared effect sizes between children that were tested in their dominant language against children that were tested in their non-dominant language, as defined by the authors. Effect sizes were larger when children were tested in their non-dominant language ($M = 0.53$, $SD = 0.90$, $range = 0-3.42$, $n = 21$) compared to their dominant language ($M = 0.35$, $SD = 0.52$, $range = 0-1.65$, $n = 23$), $Q_M(1) = 4.35$, $p = .037$. However, when inspecting Cook's distance and DFBETA values one datapoint was identified that had a relatively large effect on the outcome of the model ($g = 3.42$, standardized residual, $z = 3.00$). We therefore re-ran the model without this datapoint. Effect sizes were still slightly larger when children were tested in their non-dominant language ($M = 0.39$, $SD = 0.62$, $range = 0-1.80$, $n = 20$) compared to their dominant language ($M = 0.35$, $SD = 0.52$, $range = 0-1.65$, $n = 23$). However, this difference no longer reached significance ($Q_M(1) = 2.05$, $p = .152$). This showed that the initial significant effect was carried by the effect size that was removed.

In the second analysis, the effect of societal language was tested. The average effect size of cross-linguistic influence was larger in those situations where the language of testing was not the societal language ($M = 0.82$, $SD = 1.31$, $range = 0-7.54$, $n = 61$) compared to when it was the societal language ($M = 0.49$, $SD = 0.51$, $range = 0-2.05$,

$n = 57$), $Q_M(1) = 6.86$, $p = .009$. When inspecting Cook's distance and DFBETA values, two influential effect sizes were identified ($g = 7.54$, standardized residual, $z = 6.83$; and $g = 5.16$, standardized residual, $z = 4.82$). Without these two effect sizes, the difference in effect sizes between children tested in their non-societal language ($M = 0.64$, $SD = 0.80$, $range = 0-3.64$, $n = 59$) and in their societal languages ($M = 0.49$, $SD = 0.51$, $range = 0-2.05$, $n = 57$) was not significant but the trend was in the same direction ($Q_M(1) = 3.36$, $p = .067$). Furthermore, the estimated effect size of children tested in their non-societal and in their societal language was significantly larger than zero (non-societal language: $B = 0.70$, $SE = 0.12$, $(0.47-0.93)$, $p < .001$; societal language: $B = 0.52$, $SE = 0.12$, $(0.29-0.75)$, $p < .001$), indicating that the effect size of cross-linguistic influence was significant in the direction of the societal language into the non-societal language and in the direction of the non-societal language into the societal language.

Age

Figure 4 presents the distribution of the effect sizes by the average age of the bilingual groups by task type (107 datapoints). Two observations can be made. First, studies with younger children ($< 6;0$) in our dataset typically employed elicited production tasks to test for cross-linguistic influence. In older children, on the other hand, cross-linguistic influence was more often measured through judgement tasks. Second, the older children were, the smaller the effect of cross-linguistic influence became. This pattern was not significant, however ($Q_M(1) = 0.46$, $B = -0.003$, $SE = 0.004$, $p = .497$).

Discussion

In this study, we systematically reviewed previous research on cross-linguistic influence in bilingual children by means of a meta-analysis. Our aim was to assess the strength of cross-linguistic influence by generalizing over differences in methodology and linguistic properties. In addition, we investigated the effect of previously identified predictors of cross-linguistic influence: namely, surface overlap, language domain, language dominance, and age. A total of 26 studies met our inclusion criteria, which resulted in a total of 187 datapoints. Subsets of the available datapoints were included in the analyses testing our predictors of interest. In this section we first discuss our findings, before using them to make a number of recommendations for future studies on cross-linguistic influence.

Cross-linguistic influence: average effect size and data consistency

We assessed the presence, strength and consistency of cross-linguistic influence in previous research with bilingual children. We hypothesized that (i) there would be an overall significant effect of cross-linguistic influence, and (ii) the effect sizes of individual studies would be consistent with cross-linguistic influence. Our findings fully supported our first hypothesis and partially supported our second hypothesis.

A significant summary effect of cross-linguistic influence was observed across studies. Bilingual children's languages influence each other at the level of morpho syntax, in line with the general consensus in the literature (e.g., Serratrice, 2013). Our analyses revealed a small to moderate effect size, as reflected in a Hedges' g between 0.39 and 0.45. The moderate but not strong effect size indicates that although bilingual children's languages can influence each other, they generally behave in language-specific ways

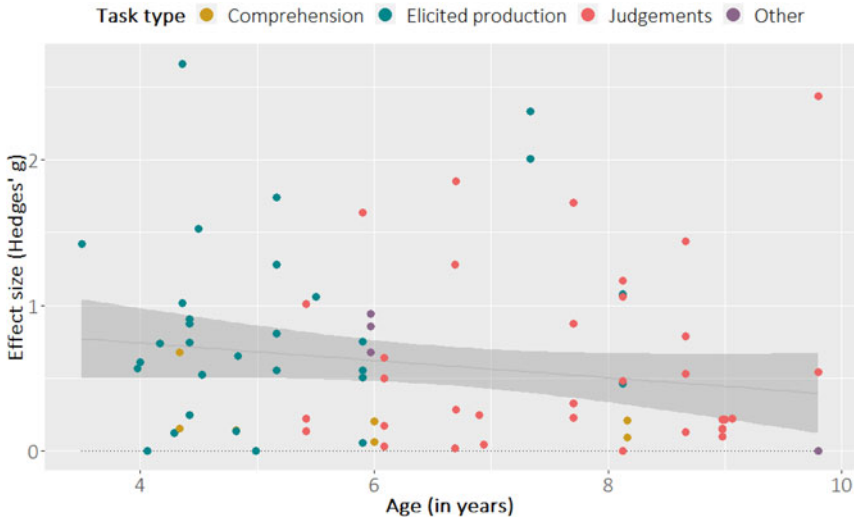


Figure 4. Effect sizes as a function of children's age (in years) by task.

similar to monolingual children (e.g., Nicoladis, 2002; Paradis & Genesee, 1996). This effect size may serve as a benchmark for future studies on cross-linguistic influence, and stimulate researchers to conduct power analyses for determining the necessary minimum sample size (e.g., Cohen, 1988).

We observed that authors did not always formulate comprehensive predictions about cross-linguistic influence. Instead, some studies focussed on certain conditions only, even when more were tested. Possibly, authors might have felt inclined to solely report significant or large effects. Indeed, the summary effects of cross-linguistic influence was slightly larger when we only took those datapoints into account for which authors made explicit predictions. There was no evidence for a publication bias in our funnel plot, however. Alternatively, authors might have focussed on conditions that offered clearest support for their theoretical perspective on cross-linguistic influence. Regardless of the reason, incomplete predictions made studies less transparent and outcomes more difficult to interpret and compare to outcomes of other studies.

Finally, most but not all datapoints in our dataset were consistent with cross-linguistic influence. Out of 113 effect sizes, 73 showed a difference between bilingual and monolingual children consistent with cross-linguistic influence. Thus, given the variety of study designs in our dataset, cross-linguistic influence can present itself regardless of the type of task set-up used or the linguistic property and language combination tested. However, 24 effect sizes went in the opposite direction and the magnitude of cross-linguistic influence varied largely across and within studies. We address this in the next sections.

Predictors of cross-linguistic influence

Surface overlap

We hypothesized that cross-linguistic influence should be stronger in situations of surface overlap versus no surface overlap. If surface overlap is a necessary condition

for cross-linguistic influence, the average effect size should be significant only in situations of surface overlap. (e.g., Hulk & Müller, 2000; Müller & Hulk, 2001). This turned out not to be the case, neither when surface overlap was coded based on authors' definitions, nor when systematically coded by us based on the adult system. The average effect size of cross-linguistic influence was not significantly different in situations of surface overlap and situations of no surface overlap.

Our analyses show that surface overlap as presently defined does not significantly affect the size of the cross-linguistic effect. However, on the basis of our results it would be inappropriate to conclude that effects of cross-linguistic influence are unaffected by ANY type of surface overlap. It is possible that when surface overlap is defined in terms of ambiguity and optionality in the child's developing system, cross-linguistic influence may still be found (e.g., Hulk & Müller, 2000; Müller & Hulk, 2001).

Take, for example, Foroodi-Nejad and Paradis' (2009) study. Their results can either be interpreted as evidence for or against the surface overlap hypothesis, depending on how surface overlap is defined. If surface overlap is based on the adult system, English constitutes a situation of no surface overlap with Persian, because English only allows right-headed compounds (whereas Persian allows both left- and right-headed compounds). If surface overlap is based on the child system, however, English might actually constitute a situation of surface overlap with Persian, because English monolingual children have been found to sometimes produce ungrammatical left-headed compounds (e.g., Foroodi-Nejad & Paradis, 2009; Nicoladis, 2002). Foroodi-Nejad and Paradis (2009) found that Persian-English children produced more left-headed compounds in English than monolingual peers. On a definition of surface overlap based on the adult system, this means that there was cross-linguistic influence in a situation of no overlap. However, on a definition based on the child system, these results constitute cross-linguistic influence in a situation of surface overlap.

Because we and most authors of the studies in our dataset defined surface overlap based on the adult system, the number of situations of surface overlap in the meta-analysis might have been underestimated. Unfortunately, we were unable to test for effects of surface overlap based on the child system as most studies provided too little information to do so. Further systematic investigation of the role of surface overlap when defined in terms of child versus the adult language system is needed.

Language domain

We hypothesized that when morphosyntax interacts with discourse pragmatics, the size of cross-linguistic influence should be stronger than in other domains. If cross-linguistic influence is only present in a domain with such an interaction (e.g., Hulk & Müller, 2000; Müller & Hulk, 2001), the average effect size of cross-linguistic influence should be significant only in this domain. This hypothesis was not borne out: there was no significant difference in effect sizes for morphosyntactic properties whose distribution was governed by discourse pragmatics compared to other morphosyntactic properties. These findings suggest that cross-linguistic influence can occur irrespective of language domain (*contra* Hulk & Müller, 2000; Müller & Hulk, 2001).

However, it proved difficult to categorise morphosyntactic properties into specific domains, as there was often no clear line between situations in which discourse pragmatics are and are not involved (e.g., Montrul, 2011; Sorace, 2011). An

alternative proposal would be to focus on computational complexity (e.g., Hopp, 2009; Sorace, 2011). Under such an account, certain morphosyntactic properties should be more sensitive to cross-linguistic influence due to their relative complexity (along the lines of Hulk and Müller's original proposal), and cross-linguistic influence could occur regardless of the language domain involved. Indeed, several studies have found evidence for the involvement of computational complexity in cross-linguistic influence (e.g., Gavarró, 2003; Strik & Pérez-Leroux, 2011).

In sum, rather than linguistic domain, computational complexity may be a more relevant predictor of cross-linguistic influence. Further investigation on this topic is needed to test this idea systematically.

Language dominance

With respect to language dominance, we hypothesized that if language dominance affects the size of cross-linguistic influence the average effect size of cross-linguistic influence would be larger from the dominant into the non-dominant language rather than the other way round. If cross-linguistic influence is from the dominant into the non-dominant language only, we predicted the effect of cross-linguistic influence to be significant in that situation only. Two analyses were conducted. We first analysed those datapoints for which the authors categorized the bilingual group as either dominant or non-dominant in the language tested. Subsequently, we operationalized language dominance in terms of the societal language. Evidence was found for the first, but not the second part of the hypothesis.

Cross-linguistic influence was stronger from children's societal language into the non-societal language than vice versa. Furthermore, the effect of cross-linguistic influence from children's non-societal language into their societal language was significantly larger than zero. In contrast, when the authors' dominance groups were analysed, no evidence for an effect of language dominance was found. Taken together, these results suggest that language dominance, as operationalized by societal language, does not predict the PRESENCE of cross-linguistic influence, but rather its STRENGTH.

The absence of an effect of dominance in the first analysis is most likely due to the differences in how authors categorized children in dominance groups. Typically, three measurements were used to assess children's dominance profile: amount of language exposure (and use), lexical proficiency, and fluency ratings by parents or teachers. Some studies combined (some of) these measurements when categorizing children into dominance groups (e.g., Foroodi-Nejad & Paradis, 2009; Pirvulescu et al., 2014). Other studies used only one of these measurements (e.g., Argyri & Sorace, 2007; Serratrice, 2007). Different measures may lead to children being assigned to different dominance groups, however (Unsworth et al., 2018).

In sum, variation WITHIN dominance groups may have masked differences *between* dominance groups in the first analysis, resulting in the absence of a significant effect of language dominance. Future studies should therefore consider testing for the effect of dominance on cross-linguistic influence by exploring different proxies for language dominance separately.

Age

With regard to age, two hypotheses were formulated: (i) if cross-linguistic influence is a developmental phenomenon, the average effect size of cross-linguistic influence should become smaller over age; (ii) if, on the other hand, cross-linguistic influence is part and

parcel of being bilingual, the average effect size of cross-linguistic influence should not differ with age.

Our results were consistent with the second hypothesis. The average effect size of cross-linguistic influence did not significantly change over age. This is in line with those previous studies that found cross-linguistic influence to remain present in older bilingual children (e.g., Argyri & Sorace, 2007; Bosch & Unsworth, 2020; Kaltsa *et al.*, 2019).

Our findings are in contrast with spontaneous production studies with very young children that attested cross-linguistic influence only during a certain phase in language development (e.g., Döpke, 1998; Hulk & Müller, 2000). This could be explained by the different modalities tested with younger and older children. In our dataset, cross-linguistic influence in older groups of bilingual children was mainly tested by judgement tasks. Possibly, these studies detected subtle effects of cross-linguistic influence that were only present in older bilingual children's judgements of sentences and not in their (spontaneous) speech production. If this is correct, cross-linguistic influence may be less strong in older bilingual children's speech production than their judgements, but this needs empirical confirmation (cf. Argyri & Sorace, 2007; Kaltsa *et al.*, 2019). It is also possible that some instances of cross-linguistic influence may be developmental in nature, whereas others are more persistent.

Two words of caution are required here. As pointed out to us by two anonymous reviewers, the effect of age on cross-linguistic influence might be more complex than it appears in the present study. First, bilingual children's age might serve as a proxy for relative exposure and as such for their language dominance. In particular, children might experience a switch in dominance from the home language to the societal language after starting school (e.g., Polinsky & Kagan, 2007). Consequently, the expected direction of cross-linguistic influence may change as children become older. Second, the relation between age and cross-linguistic influence may be modulated by the age of acquisition of the specific morpho-syntactic phenomenon in question. If cross-linguistic influence only occurs whilst children are in the process of acquiring the language property in question, then it is predicted to persist for properties that are acquired late (e.g., pronoun interpretation in languages like Italian and Greek; Papadopoulou, Peristeri, Plemenou, Marinis & Tsimpli, 2015), whereas it should be less apparent for properties that are acquired early (e.g., Verb Second in Dutch and German; Wijnen & Verrips, 1998). When the same property is acquired at different rates in different languages (e.g., gender in Greek versus gender in Dutch; Egger, Hulk & Tsimpli, 2018), this may lead to asymmetric effects of cross-linguistic influence in bilingual children acquiring those languages. By combining different morphosyntactic properties from different languages, we were unfortunately unable to disentangle effects of age from effects of age of acquisition. We encourage researchers to use the information in our dataset to conduct more in-depth analyses of age effects whilst at the same time pointing out that establishing the age of acquisition for each property in all of the relevant languages is by no means trivial. Our initial attempts to do so revealed that the necessary information was often unavailable or inconclusive.

Unexplained variation

Although some of the variance in effect sizes of cross-linguistic influence in our dataset could be explained by children's societal language, much of the variance remains unexplained, as does the observation that there were negative effect sizes. We deal with each of these issues in turn.

With respect to unexplained variance, a number of causes can be considered. First, part of the unexplained variance in effect sizes may be due to the operationalization of surface overlap and language dominance. If it had been possible to define those two constructs in a different, better way – as explained above – they might have accounted for (more) variation in the data. Our observations that the average effect size of cross-linguistic influence in situations of partial overlap and in children's dominant language was slightly but not significantly larger than in situations of no overlap and in children's non-dominant language offer support for this view.

Second, part of the unexplained variance could potentially be attributed to different types of bilingual acquisition, as pointed out to us by an anonymous reviewer. In particular, whilst some of the children in our dataset were acquiring their languages in a ONE PARENT, ONE LANGUAGE situation, others were in families where both parents spoke the minority language at home. The context in which children acquire their languages is relevant for the (cumulative) amount of input children receive (e.g., Unsworth, Brouwer, de Bree & Verhagen, 2019), which, in turn, is related to their patterns of language dominance (e.g., Unsworth et al., 2018). Consequently, average effect sizes of cross-linguistic influence might differ depending on the type of bilingual acquisition involved. Although studies in our dataset often reported at least some information about the languages spoken at home, they did not always provide the/enough relevant details. We therefore could not include the role of acquisition type in our analyses.

Third, more general effects of bilingualism could contribute to differences in performance between bilingual and monolingual children. For example, bilingual children might have performed less accurately on certain tasks compared to monolingual peers because of comparatively reduced input in their two languages or because they experienced increased processing demands having to deal with two languages instead of one (e.g., Pirvulescu et al., 2014; Sorace & Serratrice, 2009). While this latter claim remains a moot point, it is possible that general effects of bilingualism may have had a greater impact on certain morphosyntactic properties than others, and especially on those properties that require a large amount of input to be acquired or that are difficult to process. This could, in part, explain why effect sizes differed across studies. In other words, effect sizes in our dataset may not have been pure reflections of cross-linguistic influence, but may have consisted of other effects as well.

Some evidence for general bilingualism effects in bilingual children comes from a study by Sorace and colleagues (2009). They tested bilingual and monolingual children's choices of null and overt subject pronouns in Italian. They included a group of Spanish–Italian bilingual children. Spanish and Italian are both null subject languages and have similar preferences regarding subject pronoun choices (e.g., Sorace et al., 2009; but cf. Filiaci, 2010). Regardless of the overlap between languages, Sorace and colleagues found Spanish–Italian children to be less accurate in their pronoun choices than their monolingual Italian peers. Consequently, they argued that more general bilingualism effects, such as processing difficulties, affected children's pronoun choices, rather than cross-linguistic influence (also see Sorace & Serratrice, 2009 for an extensive discussion).

Fourth, all effect sizes in our dataset came from offline experiments. More recent accounts of cross-linguistic influence have suggested that cross-linguistic influence is the result of language co-activation during sentence processing (e.g., Bosch & Unsworth, 2020; Nicoladis, 2006, 2012; Serratrice, 2013, 2016). As the strength of language

co-activation may have varied from study to study – for example, due to differences in children’s language experiences – cross-linguistic influence may not always have surfaced in children’s production and offline judgements and comprehension.

Special attention should be paid to the presence of negative effect sizes. These effect sizes represented differences between bilingual and monolingual children inconsistent with cross-linguistic influence. For example, we predicted that if cross-linguistic influence was to affect French–English bilingual children’s placement of prenominal adjectives in French in Nicoladis (2006), bilingual children should be more accurate in their production of adjective-noun strings than monolingual peers. This is because English only allows for prenominal adjectives. However, bilingual children (age-matched to the monolingual children) in Nicoladis (2006) placed prenominal adjectives in French in postnominal position almost 50% of the time, versus about 10% in the French monolingual group ($g = -1.10$, $s = 0.22$). Although it could be argued that this difference between groups was a coincidence, it seems unlikely to find such a large difference between groups if cross-linguistic influence were actually present.

To account for negative effect sizes, two explanations should be considered. First, cross-linguistic influence might sometimes have resulted in a different strategy than predicted by the authors or by us. It is typically expected that cross-linguistic influence reinforces the use of a morphosyntactic structure in one of the children’s languages when it is preferred in their other language. An alternative account would be that bilingual children may sometimes try to differentiate between the morphosyntax of their languages by making their languages as different as possible (Döpke, 1998). In other words, bilingual children might adhere to canonical morpho-syntactic structures as much as possible to differentiate between languages. In the case of French, postnominal adjectives are more frequent than prenominal ones (e.g., Nicoladis, 2006). Perhaps some bilingual children in Nicoladis (2006) placed prenominal adjectives in French in postnominal position so frequently in order to contrast postnominal adjective-noun strings in French to prenominal adjective-noun strings in English. On this account, cross-linguistic influence may have led (some) bilingual children to behave in more language-specific ways than monolingual children.

It is also possible that general effects of bilingualism might explain negative effect sizes. For example, in some experiments bilingual children might have performed less accurately on a task compared to monolingual peers as a result of less input in the language tested (e.g., Pirvulescu *et al.*, 2014). This could explain why the bilingual children in Nicoladis (2006) more often incorrectly placed prenominal adjectives in French in postnominal position than monolingual children: that is, they may not have heard enough input in French to establish the prenominal position as a consistent option in that language. If a bilingualism effect were indeed responsible for children’s inconsistent behaviour with regard to cross-linguistic influence, the challenge for future studies would then be to disentangle those effects from effects of cross-linguistic influence, especially when predictions go in the same direction.

Recommendations

Facilitating reproducibility and cross-study comparisons

First of all, we recommend studies to formulate clear and testable hypotheses for each condition tested. Ideally, to make studies testing for cross-linguistic influence as

transparent as possible and less vulnerable to bias, authors should take the following steps: (i) state for all conditions tested how children's languages are different or similar; and (ii) based on this first step, state for each condition if cross-linguistic influence could manifest itself, and, importantly, *what* this cross-linguistic influence should look like when there is cross-linguistic influence and when not. Furthermore, in order to make direct comparisons across studies possible, studies should report effect sizes.

Operationalising surface overlap and language dominance

Surface overlap and language dominance should be defined and operationalized in uniform and transparent ways. With regard to surface overlap, we recommend authors to take each of the following steps: (i) describe the morphosyntactic property under study in the adult system of bilingual children's languages, at both the level of the specific context tested as well as at a more general level (for example, subjects in Greek *wh*-embedded interrogatives are always postverbal (specific context) but in other contexts they can appear preverbally as well (general context)); (ii) describe how the morphosyntactic property is acquired by monolingual and, if the relevant information is available, by bilingual children, and describe whether there is optionality during acquisition; and (iii) formulate hypotheses regarding surface overlap and indicate whether these are based on optionality in the adult language or the child language (ideally both).

With regard to language dominance, the field should strive for a standard, uniform way to define dominance. As long as this is not available, we would recommend authors to measure and report effects of amount of language exposure/use, proficiency and societal language on cross-linguistic influence separately, for example, using existing questionnaires (e.g., ALDeQ – Paradis, Emmerzael & Sorenson Duncan, 2010; BiLEC – Unsworth, 2013; PaBiQ – Tuller, 2015). This way, effects of these separate proxies of language dominance can be compared and better understood.

Cross-linguistic influence versus general effects of bilingualism

We recommend that studies differentiate effects of cross-linguistic influence from possible effects of bilingualism. For most studies in our dataset, it was impossible to determine whether effect sizes consistent with cross-linguistic influence were (partially) driven by more general effects of bilingualism as well (cf. Pirvulescu et al., 2014; Serratrice et al., 2009, 2012; Sorace et al., 2009). We therefore propose that future studies include, where possible, an appropriate bilingual control group (e.g., Kaltsa et al., 2019; Serratrice et al., 2009, 2012; Sorace et al., 2009). Crucially, without this bilingual control group, it may be impossible to determine whether differences between a bilingual and monolingual group should be attributed to cross-linguistic influence or to a more general bilingualism effect (for similar discussion concerning adult second language learners, see Jarvis, 2000).

We do realize that for practical reasons it is not always possible to add a control group. In these situations, we recommend authors to consider the introduction of multiple within-experiment conditions that test the same cross-linguistic effects in different ways, and/or the inclusion of matched control-conditions in which only general bilingual effects would be expected (e.g., complete-overlap conditions).

Effect sizes from these studies could then be used to calculate a more precise average effect size of cross-linguistic influence.

Sample size and power

Ideally, future studies should consider the minimum sample size of children necessary to obtain a significant effect of cross-linguistic influence. If the true effect size of cross-linguistic influence is 0.39, then a sample size of at least 82 children would be necessary in the bilingual and monolingual control group to detect this effect (for an alpha level of .05 and a beta level of .80). If the true effect size is 0.45 a minimum sample size of 62 children per group would be necessary (calculations were performed with G*power; Faul, Erdfelder, Lang & Buchner, 2007). This means that with just one exception (Meir *et al.*, 2017), all the studies in our dataset will likely have been underpowered. In fact, the vast majority of studies did not even test half of the participants required. We do realize that increasing sample sizes is easier said than done, especially given the relative scarcity of certain bilingual populations and the labour intensity of the data collection process. One solution to the power problem would be for researchers to collaborate when possible (Brybaert, 2019).

Apart from testing more participants, researchers could aim to increase the sensitivity of their studies by decreasing error variance as much as possible. For example, by keeping background variables, such as age, proficiency, and amount of exposure – or, if not possible, type of acquisition – as constant as possible between bilingual children and by increasing the numbers of items tested (Brybaert, 2019; see also Quené, 2010 for a further discussion how to increase the sensitivity of a study). For example, in 19 studies in our dataset at least some of the reported group means were based on less than 10 items per condition and in four studies there were even less than five items. This might have resulted in less precise outcomes – and therefore decreased power – compared to studies with more items per condition. Furthermore, many studies in our dataset reported rather broad language proficiency and/or exposures ranges for bilingual children (e.g., Cuza & Pérez-Tattam, 2016; Foroodi-Nejad & Paradis, 2009; Serratrice *et al.*, 2009). It is possible that children with very different language profiles show different effects of cross-linguistic influence from other children, especially given our finding that language dominance affects the strength of cross-linguistic influence. Combining results from children with very different backgrounds might therefore increase the noise in the data, decreasing the likelihood of differences between bilingual and monolingual scores reaching significance. Moreover, one solution frequently adopted by authors of splitting children into different groups decreases the sample size, again resulting in a loss of power. As an alternative, authors could strive to select bilingual children with as similar linguistic background as possible to obtain more precise group effects.

Finally, our estimation of a minimum sample size of 62 to 82 children per group is based on the average effect size of cross-linguistic influence from studies for which it is unclear to what extent a more general effect of bilingualism was at play. Other factors might have affected the effect size of cross-linguistic influence that we were unable to test for in this meta-analysis and hence the effect size reported here may be an over- or underestimation. In the latter case, smaller minimum sample sizes would be required for a properly powered study. Future studies following our recommendations are necessary to clarify this issue further.

Understudied areas of cross-linguistic influence

Finally, we recommend conducting additional studies on cross-linguistic influence in children's comprehension. The majority of studies in our dataset were concerned with elicited production or judgement tasks and only a few studies concerned

comprehension (Nicoladis, 2003; Serratrice, 2007; Syrett, Lingwall, Perez-Cortes, Austin, Sánchez, Baker, Germak & Arias-Amaya, 2017; van Koert, Koeneman, Hulk & Weerman, 2016). It is therefore unclear whether the average effect sizes attested in our meta-analysis apply to cross-linguistic influence in comprehension as well.

Furthermore, all studies in our dataset focussed on cross-linguistic influence in children's offline production, judgements and comprehension. Until now, virtually no studies have focused on cross-linguistic influence during real-time morphosyntactic processing (cf. Lemmerth & Hopp, 2019). This, too, might have resulted in an underestimation of cross-linguistic influence attested in bilingual children. More online data are necessary to explore more subtle effects of cross-linguistic influence.

Conclusion

This meta-analysis is the first study to systematically assess the effect size of cross-linguistic influence in bilingual children and effects of surface overlap, language domain, language dominance and age. Overall, there was a significant effect of cross-linguistic influence across studies and its average effect size was small to moderate. Furthermore, the results of most of the studies were consistent with cross-linguistic influence. Cross-linguistic influence was stronger from children's societal language into their non-societal language than vice versa. No effects were found for surface overlap – either as defined by the authors of the studies or based on the adult language system only – language domain, language dominance as operationalized by the authors of the studies, or age. These findings suggest that cross-linguistic influence is part and parcel of being bilingual and can manifest itself in various linguistic contexts. At the same time, our meta-analysis also shows that more systematic and standardized studies of cross-linguistic influence are necessary to fully understand this aspect of bilingual language development and use. This especially holds for the formulation of hypotheses about cross-linguistic influence and the operationalization of surface overlap and language dominance. We hope that the recommendations given here will serve as an impetus for the field to move towards a more standardized and unified way of testing for cross-linguistic influence and its predictors.

Acknowledgements. We are very grateful to all the authors who answered our questions about their studies and who dug into their archives to provide us with their data. We would also like to thank two anonymous reviewers for their constructive comments. This work was supported by a grant from the Netherlands Organisation for Scientific Research (NWO; project number 23000408) to the first author and (NWO; project number 23000407) to the last author.

Supplementary Material. For supplementary material accompanying this paper, visit <https://doi.org/10.1017/S0305000921000337>.

References

References marked with an asterisk indicate studies included in the meta-analysis.

- *Argyri, E., & Sorace, A. (2007). Crosslinguistic influence and language dominance in older bilingual children. *Bilingualism: Language and Cognition*, 10(1), 79–99. <https://doi.org/10.1017/S1366728906002835>
- Austin, J. (2007). Grammatical interference and the acquisition of ergative case in bilingual children learning Basque and Spanish. *Bilingualism: Language and Cognition*, 10(3), 315–331. <https://doi.org/10.1017/S1366728907003094>
- Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C., & Cristia, A. (2018). Promoting replicability in developmental research through meta-analyses: Insights from language acquisition research. *Child Development*, 89(6), 1996–2009. <https://doi.org/10.1111/cdev.13079>

- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons. <https://doi.org/10.1002/9780470743386>
- *Bosch, J. E., & Unsworth, S. (2020). Cross-linguistic influence in word order: effects of age, dominance and surface overlap. *Linguistic Approaches to Bilingualism*, 1–34. <https://doi.org/10.1075/lab.18103.bos>
- Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, 2(1), 1–38. <https://doi.org/10.5334/joc.72>
- Carminati, M. N. (2002). *The processing of Italian subject pronouns*. [Doctoral dissertation, University of Massachusetts]. Proquest. <https://scholarworks.umass.edu/dissertations/AAI3039345>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203771587>
- *Cuza, A., & Pérez-Tattam, R. (2016). Grammatical gender selection and phrasal word order in child heritage Spanish: A feature re-assembly approach. *Bilingualism: Language and Cognition*, 19(1), 50–68. <https://doi.org/https://doi.org/10.1017/S1366728914000893>
- Daskalaki, E., Chondrogianni, V., Blom, E., Argyri, F., & Paradis, J. (2019). Input effects across domains: The case of Greek subjects in child heritage language. *Second Language Research*, 35(3), 421–445. <https://doi.org/10.1177/0267658318787231>
- de Houwer, A. (1990). *The acquisition of two languages from birth*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511519789>
- Döpke, S. (1998). Competing language structures: the acquisition of verb placement by bilingual German-English children. *Journal of Child Language*, 25(3), 555–584. <https://doi.org/10.1017/S0305000998003584>
- *Egger, E., Hulk, A., & Tsimpli, I. M. (2018). Crosslinguistic influence in the discovery of gender: The case of Greek-Dutch bilingual children. *Bilingualism: Language and Cognition*, 21(4), 694–709. <https://doi.org/10.1017/S1366728917000207>
- *Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). GPower 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Fernández-Castilla, B., Jamshidi, L., Declercq, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2020). The application of meta-analytic (multi-level) models with multiple random effects: A systematic review. *Behavior Research Methods*, 52, 2031–2052. <https://doi.org/10.3758/s13428-020-01373-9>
- Fernández-Castilla, B., Maes, M., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2019). A demonstration and evaluation of the use of cross-classified random-effects models for meta-analysis. *Behavior Research Methods*, 51(3), 1286–1304. <https://doi.org/10.3758/s13428-018-1063-2>
- Filiaci, F. (2010). Null and overt subject biases in Spanish and Italian: A cross-linguistic comparison. In C. Borgonovo, M. Español-Echevarría, & P. Prévost (Eds.), *Selected proceedings of the 12th Hispanic Linguistics Symposium* (pp. 171–182). Cascadilla Proceedings project.
- *Foroodi-Nejad, F., & Paradis, J. (2009). Crosslinguistic transfer in the acquisition of compound words in Persian-English bilinguals. *Bilingualism: Language and Cognition*, 12(4), 411–427. <https://doi.org/10.1017/S1366728909990241>
- *Gathercole, V. C. M., Laporte, N., & Thomas, E. M. (2005). Differentiation, carry-over, and the distributed characteristic in bilinguals: Structural “mixing” of the two languages? In J. Cohen, K. T. McAlister, K. Rolstad, & J. MacSwan (Eds.), *Proceedings of the 4th International Symposium on Bilingualism (ISB4)* (pp. 838–851). Cascadilla Press.
- Gavarró, A. (2003). Economy and word order patterns in bilingual English-Dutch acquisition. *Bilingualism: Language and Cognition*, 6(1), 69–79. <https://doi.org/10.1017/S1366728903001044>
- Genesee, F., Paradis, J., & Crago, M. B. (2004). *Dual language development & disorders: A handbook on bilingualism & second language learning*. Paul Brookes.
- Grosjean, F. (1982). *Life with two languages: An introduction to bilingualism*. Harvard University Press.
- Haznedar, B. (2007). Crosslinguistic influence in Turkish-English bilingual first language acquisition: The overuse of subjects in Turkish. In A. Belikova, L. Meroni, & M. Umeda (Eds.), *Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition North America (GALANA)* (pp. 124–134). Cascadilla Proceedings Project.

- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107–128. <https://doi.org/10.2307/1164588>
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.
- *Hervé, C., & Lawyer, L. (unpublished manuscript). *Bilinguals' sensitivity to specificity and genericity: Evidence from implicit and explicit knowledge*. 1–52.
- Hervé, C., Serratrice, L., & Corley, M. (2016). Dislocations in French–English bilingual children: An elicitation study. *Bilingualism: Language and Cognition*, 19(5), 987–1000. <https://doi.org/10.1017/S1366728915000401>
- Hopp, H. (2009). The syntax–discourse interface in near-native L2 acquisition: Off-line and on-line performance. *Bilingualism: Language and Cognition*, 12(4), 463–483. <https://doi.org/10.1017/S1366728909990253>
- Hoyt, W. T., & Del Re, A. C. (2018). Effect size calculation in meta-analyses of psychotherapy outcome research. *Psychotherapy Research*, 28(3), 379–388. <https://doi.org/10.1080/10503307.2017.1405171>
- Hulk, A., & Müller, N. (2000). Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: Language and Cognition*, 3(3), 227–244. <https://doi.org/10.1017/S1366728900000353>
- Jarvis, S. (2000). Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon. *Language Learning*, 50(2), 245–309. <https://doi.org/10.1111/0023-8333.00118>
- Kaltsa, M., Tsimpli, I. M., & Argyri, F. (2019). The development of gender assignment and agreement in English-Greek and German-Greek bilingual children. *Linguistic Approaches to Bilingualism*, 9(2), 253–288. <https://doi.org/10.1075/lab.16033.kal>
- *Kidd, E., Chan, A., & Chiu, J. (2015). Cross-linguistic influence in simultaneous Cantonese–English bilingual children's comprehension of relative clauses. *Bilingualism: Language and Cognition*, 18(3), 438–452. <https://doi.org/10.1017/S1366728914000649>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 1–12. <https://doi.org/10.3389/fpsyg.2013.00863>
- Lemmerth, N., & Hopp, H. (2019). Gender processing in simultaneous and successive bilingual children: Cross-linguistic lexical and syntactic influences. *Language Acquisition*, 26(1), 21–45. <https://doi.org/10.1080/10489223.2017.1391815>
- McLaughlin, B. (1978). *Second language acquisition in childhood*. Erlbaum.
- *Meir, N., Walters, J., & Armon-Lotem, S. (2017). Bi-directional cross-linguistic influence in bilingual Russian–Hebrew children. *Linguistic Approaches to Bilingualism*, 7(5), 514–553. <https://doi.org/10.1075/lab.15007.mei>
- Meisel, J. M. (1989). Early differentiation of languages in bilingual children. In K. Hyltenstam & L. Obler (Eds.), *Bilingualism across the lifespan: Aspects of acquisition, maturity and loss* (pp. 13–40). Cambridge University Press. <https://doi.org/10.1017/CBO9780511611780.003>
- Meisel, J. M. (2007). On autonomous syntactic development in multiple first language acquisition. In H. Caunt-Nulton, S. Kulatilake, & I.-H. Woo (Eds.), *Proceedings of the 31th annual Boston University Conference on Language Development (BUCLD 31)* (pp. 26–45). Cascadilla Press.
- *Meroni, L., Smeets, L., & Unsworth, S. (2017). Cross-linguistic influence in scope ambiguity: Evidence for acceleration. In E. Blom, L. Cornips, & J. Schaeffer (Eds.), *Cross-linguistic influence in Bilingualism: In honor of Aafke Hulk* (pp. 181–206). John Benjamins. <https://doi.org/10.1075/sibil.52.09mer>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), 1–6. <https://doi.org/10.1371/journal.pmed.1000097>
- Montrul, S. (2011). Multiple interfaces and incomplete acquisition. *Lingua*, 121(4), 591–604. <https://doi.org/10.1016/j.lingua.2010.05.006>
- Müller, N., & Hulk, A. (2001). Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism: Language and Cognition*, 4(1), 1–21. <https://doi.org/10.1017/S1366728901000116>
- *Mykhaylyk, R., & Ytterstad, E. (2017). Directionality of cross-linguistic influence: Which referring choices do bilingual Ukrainian–English children make? *International Journal of Bilingualism*, 21(1), 99–121. <https://doi.org/10.1177/1367006915603824>
- *Nicoladis, E. (2002). What's the difference between 'toilet paper' and 'paper toilet'? French–English bilingual children's crosslinguistic transfer in compound nouns. *Journal of Child Language*, 29(4), 843–863. <https://doi.org/10.1017/S0305000902005366>
- *Nicoladis, E. (2003). Cross-linguistic transfer in deverbal compounds of preschool bilingual children. *Bilingualism: Language and Cognition*, 6(1), 17–31. <https://doi.org/10.1017/S1366728903001019>

- *Nicoladis, E. (2006). Cross-linguistic transfer in adjective–noun strings by preschool bilingual children. *Bilingualism: Language and Cognition*, 9(1), 15–32. <https://doi.org/10.1017/S136672890500235X>
- *Nicoladis, E. (2012). Cross-linguistic influence in French–English bilingual children’s possessive constructions. *Bilingualism: Language and Cognition*, 15(2), 320–328. <https://doi.org/10.1017/S1366728911000101>
- *Nicoladis, E., & Gavrilá, A. (2015). Cross-linguistic influence in Welsh–English bilingual children’s adjectival constructions. *Journal of Child Language*, 42(4), 903–916. <https://doi.org/https://doi.org/10.1017/S0305000914000440>
- Papadopoulou, D., Peristeri, E., Plemenou, E., Marinis, T., & Tsimpli, I. (2015). Pronoun ambiguity resolution in Greek: Evidence from monolingual adults and children. *Lingua*, 155, 98–120. <https://doi.org/10.1016/j.lingua.2014.09.006>
- Paradis, J., Emmerzael, K., & Sorenson Duncan, T. (2010). Assessment of English language learners: Using parent report on first language development. *Journal of Communication Disorders*, 43(6), 474–497. <https://doi.org/10.1016/j.jcomdis.2010.01.002>
- Paradis, J., & Genesee, F. (1996). Syntactic acquisition in bilingual children. *Studies in Second Language Acquisition*, 18(1), 1–25. <https://doi.org/10.1017/S0272263100014662>
- Paradis, J., & Navarro, S. (2003). Subject realization and crosslinguistic interference in the bilingual acquisition of Spanish and English: What is the role of the input? *Journal of Child Language*, 30(2), 371–393. <https://doi.org/10.1017/S0305000903005609>
- *Pirvulescu, M., Pérez-Leroux, A.-T., Roberge, Y., Strik, N., & Thomas, D. (2014). Bilingual effects: Exploring object omission in pronominal languages. *Bilingualism: Language and Cognition*, 17(3), 495–510. <https://doi.org/10.1017/S1366728913000631>
- Polinsky, M., & Kagan, O. (2007). Heritage languages: In the ‘wild’ and in the classroom. *Language and Linguistics Compass*, 1(5), 368–395. <https://doi.org/10.1111/j.1749-818X.2007.00022.x>
- Quené, H. (2010). How to design and analyze language acquisition studies. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 269–284). John Benjamins. <https://doi.org/10.1075/llt.27.15que>
- R Core Team. (2020). *R: A language and environment for statistical computing* (version 4.0.3) [Computer software]. Retrieved from <https://www.r-project.org/>
- Rodina, Y., Kupisch, T., Meir, N., Mitrofanova, N., Urek, O., & Westergaard, M. (2020). Internal and external factors in heritage language acquisition: Evidence from heritage Russian in Israel, Germany, Norway, Latvia and the United Kingdom. *Frontiers in Education*, 5, 1–17. <https://doi.org/10.3389/educ.2020.00020>
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. John Wiley & Sons. <https://doi.org/10.1002/0470870168>
- Schmitz, K., Patuto, M., & Müller, N. (2012). The null-subject parameter at the interface between syntax and pragmatics: Evidence from bilingual German–Italian, German–French and Italian–French children. *First Language*, 32(1–2), 205–238. <https://doi.org/10.1177/0142723711403880>
- Schwartz, M., Minkov, M., Dieser, E., Protassova, E., Moin, V., & Polinsky, M. (2015). Acquisition of Russian gender agreement by monolingual and bilingual children. *International Journal of Bilingualism*, 19(6), 726–752. <https://doi.org/10.1177/1367006914544989>
- *Serratrice, L. (2007). Cross-linguistic influence in the interpretation of anaphoric and cataphoric pronouns in English–Italian bilingual children. *Bilingualism: Language and Cognition*, 10(3), 225–238. <https://doi.org/10.1017/S1366728907003045>
- Serratrice, L. (2013). Cross-linguistic influence in bilingual development: Determinants and mechanisms. *Linguistic Approaches to Bilingualism*, 3(1), 3–25. <https://doi.org/10.1075/lab.3.1.01ser>
- Serratrice, L. (2016). Cross-linguistic influence, cross-linguistic priming and the nature of shared syntactic structures. *Linguistic Approaches to Bilingualism*, 6(6), 822–827. <https://doi.org/10.1075/lab.6.6.15ser>
- *Serratrice, L., Sorace, A., Filiaci, F., & Baldo, M. (2009). Bilingual children’s sensitivity to specificity and genericity: Evidence from metalinguistic awareness. *Bilingualism: Language and Cognition*, 12(2), 239–257. <https://doi.org/10.1017/S1366728909004027>
- *Serratrice, L., Sorace, A., Filiaci, F., & Baldo, M. (2012). Pronominal objects in English–Italian and Spanish–Italian bilingual children. *Applied Psycholinguistics*, 33(4), 725–751. <https://doi.org/10.1017/S0142716411000543>
- Serratrice, L., Sorace, A., & Paoli, S. (2004). Crosslinguistic influence at the syntax–pragmatics interface: Subjects and objects in English–Italian bilingual and monolingual acquisition. *Bilingualism: Language and Cognition*, 7(3), 183–205. <https://doi.org/10.1017/S1366728904001610>

- Sorace, A.** (2011). Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism*, 1(1), 1–33. <https://doi.org/10.1075/lab.1.1.01sor>
- Sorace, A., & Serratrice, L.** (2009). Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism*, 13(2), 195–210. <https://doi.org/10.1177/1367006909339810>
- ***Sorace, A., Serratrice, L., Filiaci, F., & Baldo, M.** (2009). Discourse conditions on subject pronoun realization: Testing the linguistic intuitions of older bilingual children. *Lingua*, 119(3), 460–477. <https://doi.org/10.1016/j.lingua.2008.09.008>
- Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. T.** (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, 343(d4002), 1–8. <https://doi.org/10.1136/bmj.d4002>
- ***Strik, N., & Pérez-Leroux, A. T.** (2011). Jij doe wat girafe?: Wh-movement and inversion in Dutch-French bilingual children. *Linguistic Approaches to Bilingualism*, 1(2), 175–205. <https://doi.org/10.1075/lab.1.2.03str>
- ***Syrett, K., Lingwall, A., Perez-Cortes, S., Austin, J., Sánchez, L., Baker, H., Germak, C., & Arias-Amaya, A.** (2017). Differences between Spanish monolingual and Spanish-English bilingual children in their calculation of entailment-based scalar implicatures. *Glossa: A Journal of General Linguistics*, 2(1), 1–19. <https://doi.org/10.5334/gjgl.76>
- Tuller, L.** (2015). Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 301–330). Multilingual Matters. <https://doi.org/10.21832/9781783093137-013>
- Unsworth, S.** (2003). Testing Hulk & Müller (2000) on crosslinguistic influence: Root Infinitives in a bilingual German/English child. *Bilingualism: Language and Cognition*, 6(2), 143–158. <https://doi.org/10.1017/S1366728903001093>
- Unsworth, S.** (2012). Testing for crosslinguistic influence and exposure effects in the bilingual acquisition of specific indefinite objects. In A. K. Biller, E. Y. Chung, & A. E. Kimball (Eds.), *Proceedings of the 36th annual Boston University Conference on Language Development (BUCLD 36)* (pp. 633–645). Cascadilla Press.
- Unsworth, S.** (2013). Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition*, 16(1), 86–110. <https://doi.org/10.1017/S1366728912000284>
- Unsworth, S., Brouwer, S., de Bree, E., & Verhagen, J.** (2019). Predicting bilingual preschoolers’ patterns of language development: Degree of non-native input matters. *Applied Psycholinguistics*, 40(5), 1189–1219. <https://doi.org/10.1017/S0142716419000225>
- Unsworth, S., Chondrogianni, V., & Skarabela, B.** (2018). Experiential measures can be used as a proxy for language dominance in bilingual language acquisition research. *Frontiers in Psychology*, 9, 1–15. <https://doi.org/10.3389/fpsyg.2018.01809>
- van Dijk, C.N., van Wonderen, Koutamanis, E., Kootstra, G.J., & Unsworth, S.** (2021). Effect sizes of experimental studies on cross-linguistic influence in bilingual children. *DANS*. <https://doi.org/10.17026/dans-z6a-yw53>.
- ***van Koert, M., Koeman, O., Hulk, A., & Weerman, F.** (2016). Revisiting binding in Turkish-Dutch bilingual children: Comprehension and production. In D. Stringer, J. Garrett, B. Halloran, & S. Mossman (Eds.), *Proceedings of the 13th Generative Approaches to Second Language Acquisition Conference (GASLA 2015)* (pp. 80–94). Cascadilla Proceedings Project.
- Viechtbauer, W.** (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
- Wijnen, F., & Verrips, M.** (1998). The acquisition of Dutch syntax. In S. Gillis & A. De Houwer (Eds.), *The acquisition of Dutch* (pp. 233–299). <https://doi.org/10.1075/pbns.52>
- Yip, V., & Matthews, S.** (2000). Syntactic transfer in a Cantonese–English bilingual child. *Bilingualism: Language and Cognition*, 3(3), 193–208. <https://doi.org/10.1017/S13667289000033X>

Cite this article: Van Dijk C, Van Wonderen E, Koutamanis E, Kootstra GJ, Dijkstra T, Unsworth S (2021). Cross-linguistic influence in simultaneous and early sequential bilingual children: a meta-analysis. *Journal of Child Language* 1–33. <https://doi.org/10.1017/S0305000921000337>