Chapter 5 - General Discussion

Caroline Floccia, Thomas D. Sambrook, Claire Delle Luche, Rosa Kwok, Jeremy Goslin, Laurence White, Allegra Cattani, Emily Sullivan, Kirsten Abbot-Smith, Andrea Krott, Debbie Mills, Caroline Rowland, Judit Gervain, Kim Plunkett

Corresponding author: Caroline Floccia, caroline.floccia@plymouth.ac.uk

Ours is the first study to directly measure the relative contribution of linguistic distance to the acquisition of a bilingual toddler's two languages, which was the aim of Study 1. To address this question we tested 372 24-month-olds learning British English and one of 13 target Additional Languages. We found that a higher phonological overlap between these Additional Languages and British English led to higher levels of Additional Language CDI vocabulary production. Similar effects in comprehension were found for our other measures of linguistic distance, namely degree of similar in morphological complexity and word order. Importantly, linguistic distance contributed unique variance even when other key factors (proportion of English in child-directed input, proportion of English in parental overheard speech and gender) were entered into the same model.

Ours is also the first study to develop bilingual norms specified for the proportion of a particular language that a child hears in the input, which was the objective of Study 2. To establish the reliability of our norms, we applied the model developed with the 372 24-month-olds learning British English and one of 13 target Additional Languages, to a cohort of 58 24-month-olds learning British English and an Additional Language which was not part of our 13 target languages. We found that the English vocabulary scores derived from the model were highly predictive of the vocabulary scores of these 58 children, showing a strong validity of our model for estimating word knowledge in any bilingual toddler growing up in the UK, which was the aim of Study 3. This demonstrates the feasibility of assessing bilingual toddlers in the majority language only, providing professionals working with young children with a practical

solution to a long-standing problem in societies where bilinguals come from heterogeneous backgrounds.

Study 1: Linguistic Distance and the Bilingual Lexicon

To explore the effect of linguistic distance on vocabulary outcome, we were faced with the problem of disentangling variance due to linguistic distance from that due to cultural diversity amongst language communities. In our initial analyses we treated our 13 target language pairs as a random factor. This provided an excellent fit between predicted and observed values from bilinguals learning an Additional Language outside of our 13 target languages, indicating that this factor should be included in any future study using heterogeneous sampling of bilinguals. In our second analysis we replaced the random effect of language with a linear predictor of linguistic distance, as measured by the phonological overlap between English words and their translation equivalents in the Additional Language. We found that children's production of Additional Language words was improved when this language was phonologically close to English (such as Dutch, Welsh, German) as compared to more distant languages (such as Cantonese, Polish and Greek). Testing other measures of linguistic distance, we found that children have a receptive vocabulary boost in their Additional Language if they learn a language with the same word order typology as British English (such as Polish or Portuguese) and/or a morphologically close language (such as Cantonese or Mandarin).

Phonological overlap distance. In the literature the possibility that bilinguals do not face the same learning challenges depending on the language pairs being acquired is rarely acknowledged (but see Argyri & Sorace, 2007). Here we established that linguistic distance between a bilingual child's languages shapes their word learning at the age of two. This finding is perhaps less surprising given the recent growing evidence from monolingual research showing that young learners follow slightly different paths in speech perception (e.g., Mani &

Plunkett, 2010; Nazzi, 2005), segmentation (e.g., Höhle et al., 2009) and lexical growth (e.g., Bleses et al., 2008) depending on the language they acquire. It is also supported by research examining adult bilingual lexical access which has investigated the effect of lexical overlap of translation equivalents on lexical access (e.g., Costa, Caramazza, & Sebastián-Gallés, 2000; Strijkers, Costa, & Thierry, 2009). Cognates, translation equivalents with form overlap, such as *bed* and *Bett* in German, are known to provide an advantage to production, similar to the linear measure of phonological overlap used in the current study. In word production, cognates are produced faster (e.g., Costa et al., 2000; Hoshino & Kroll, 2008), elicit a different brain signature (e.g., Strijkers et al., 2009), and produce higher levels of activation in priming tasks (e.g., Colomé & Miozzo, 2010) than non-cognates. The cognate advantage in picture naming is also found in young proficient bilinguals as early as 4 years (Sheng, Lam, Cruz, & Fulton, 2016; see also Poarch & van Hell, 2012).

To explain this advantage it has been proposed that cognates have different representations than non-cognates in the lexicon, perhaps because of a larger conceptual overlap (van Hell & De Groot, 1998), a shared morphological representation (Sánchez-Casas & García-Albea, 2005), or because they might have been learned earlier (Costa, Pannunzi, Deco, & Pickering, 2016). Alternatively, it has been proposed that the cognate advantage is the by-product of the dynamic interactions between the lexical and the phonological (and orthographical) levels of processing within the lexicon (Costa, Sanesteban, & Caño, 2005). Altogether, these findings with cognates show good equivalence with our own findings, where phonological overlap between translation equivalents increased Additional Language expressive vocabulary. This finding also provides support to the proposal that the cognate advantage is due to cognates being acquired before non-cognates in early childhood (Costa et al., 2016), leading to an ease of processing later in life.

Finally, the cognate advantage is an emerging property of recent computational models of the bilingual lexicon (BLINCS: Shook & Marian, 2013; DEV-LEX II: Zhao & Li, 2010): the activation of a concept in a common semantic lexicon during word production would generate parallel top-down activation of phonological representations from lexical representations in both languages. Any segments that overlapped between language representations would receive top-down activation from both of the languages lexical representations. This would provide an advantage to cognate or more overlapped translation equivalents, as they would become activated faster than those with less or non-overlapped representations (see also Costa et al., 2005).

One aspect of our results worth noting is that, although we found an increase in Additional Language vocabulary with phonological overlap in production, we found no effect in comprehension. This disparity also has some parallels in prior literature, where the evidence for a cognate advantage in bilingual spoken word recognition is mixed (see a review in Lagrou, Hartsuiker, & Duyck, 2011; see the review by Sánchez-Casas & García-Albea, 2005, for the cognate effect in *visual* word recognition). Although there is evidence for non-selective lexical access in word recognition as is found in production (Lagrou et al., 2011), that is, for online activation of words in both language, Shook, Goldrick, Engstler and Marian (2015) did not find any cognate advantage in a word recognition task in spoken English sentences in German-English bilinguals. When using an eye-tracking word recognition task for sets of pictures, Blumenfeld and Marian (2007) reported a cognate advantage dependent on the listener's proficiency in the target language. That is, whereas a cognate and non-cognate equally activate words in a highly proficient language, only cognates boost the activation of words in a less proficient language. In development all prior studies of cognate advantage in spoken word recognition have been with Spanish-English bilinguals. Again, findings are mixed, with some studies not finding any advantage in receptive vocabulary recognition (in first graders: Umbel,

Pearson, Fernandez, & Oller, 1992; first to sixth graders: Umbel & Oller, 1994), while others did find an advantage (fifth and sixth graders: Cunningham & Graham, 2000; 8 to 13 years old: Kelley & Kohnert, 2012; kindergarten and first graders: Pérez, Peña, & Bedore, 2010). Kelley and Kohnert (2012) actually directly compared the cognate effect in production and comprehension using standardised tests, and found a very small advantage in comprehension, and a small to medium effect in production. In comprehension a large proportion of the variance in the cognate advantage was due to age, with older children showing a greater advantage than younger ones.

In sum, our findings are the first to demonstrate that phonological overlap at the lexical level appears to boost the acquisition of expressive vocabulary in bilingual toddlers, which fits nicely with past demonstrations of a cognate advantage in older children and adults.

Syntactic and morphological distance. Our findings of a facilitatory effect of word order typology distance on receptive vocabulary can be explained in three, non-exclusive ways. First, children learning two languages with the same word order, such as British English and Polish, can probably use similar phrase-level prosodic cues for segmentation of syntactic constituents: indeed VO languages (such as English and French) primarily use duration to express prosodic prominence of a content word as compared to a functor, while OV languages (such as Bengali and Hindi) tend to use pitch/intensity (Nespor et al., 2008). The search for similar cues in the speech signal would lead to a single mechanism for prosodic-driven segmentation, boosting the retrieval of words and the assignment of syntactic categories. An additional explanation of this facilitatory effect is that, instead of activating both languages, bilinguals transfer the structure of their native language processing to that of the new language (Costa et al., 2016). Although this proposal applies to lexical processing in sequential bilinguals, the idea that a processing structure for one language can be ‘carried over’ to the other language also has currency for word order computation. Finally, some studies have

proposed that word order variations amongst the world's languages are constrained by computational or learnability limitations, leading to a proliferation of easy-to-learn orders (e.g., Ferrer-i-Cancho, 2015; Lupyan & Christiansen, 2002). In this perspective, pairs of languages that differ on the word order dimension will add to the learnability issue by adding computational complexity.

In addition to word order typology we also found a facilitative effect of morphological similarity, with children learning Additional Languages morphologically close to English, such as isolating Cantonese or Mandarin, having better Additional Language comprehension vocabulary than those with a distant Additional Language, such as the agglutinative Bengali. In monolingual research, it has been argued that complex morphological systems may hinder language acquisition (e.g., Slobin, 1973) - although complex systems can be learnt quickly provided that morphological rules are regular and obligatory (e.g., Devescovi et al., 2005; Kim, McGregor, & Thompson, 2000). Our findings suggest that bilinguals learning languages with similar morphology may benefit from the training of cognitive resources engaged in supporting the learning of one kind of morphology over another, for example, memory for isolating languages, versus rule-based learning for synthetic languages, as suggested by Fortescue and Lennert Olsen (1992). Some aspects of this cognitive training might be driving the bilingual cognitive advantage (Bialystok, 2009), raising the interesting possibility that linguistic distance could be partially responsible for the elusive nature of this cognitive advantage (Paap & Greenberg, 2013).

On a final note, when considering these findings it should be noted that language distance was found to modulate Additional Language scores only, with English scores being more resistant to facilitatory effects from the Additional Language proximity. We interpret this as showing that English acquisition benefits from the overarching effect of the English-speaking environment, whereas the Additional Language acquisition relies, in most cases,

solely on one or two parents' input. While measures of exposure do not quantify the weight of everyday social interactions (shopping, visit to the doctor, media exposure, etc), a telling outcome that can be derived from Figure 1 is that in order for a bilingual child from our corpus to know an equal proportion of words in English and Additional Language, her/his exposure to English as measured by the Plymouth Language Exposure Questionnaire only needs to be at 30% of English in comprehension (they would then know 23 out the 30 common words in English and the Additional Language) and 20% in production (they would then produce 14 out of the 30 common words in English and the Additional Language). These low English exposure values clearly fail to capture the overwhelming influence of the language of the surrounding community, and point to the relative vulnerability of the growth of the Additional Language.

In this study we have focussed on three measures of linguistic distance, phonological overlap of the child lexicon, syntax typology and morphological complexity, that reflect the core aspects of language development in the second year of life. However, 'linguistic distance' is fundamentally complex, with a multidimensional representation that goes beyond the factors examined thus far. Our database provides an opportunity to continue the exploration of how additional measures of linguistic distance can account for variations in bilingual vocabulary, for example, through examination of the effect of prosody on early word development (e.g., rhythmic families, final word lengthening, etc; White et al., in prep), or the effect of cross-linguistic differences in infant-directed speech styles (e.g., Fernald et al., 1989).

In conclusion, the finding that linguistic distance shapes vocabulary knowledge in the early years strongly supports the idea of independent (Genesee et al., 1995), yet interfering language systems from the onset of development (DeAnda et al., 2016, Von Holzen & Mani, 2012). It is unknown at this stage whether the facilitation effects we found for phonologically or structurally close languages are due to the sharing of information processing mechanisms, or to the duplication of processing structures from one language to the next, following the idea

of a transfer proposed by Costa et al. (2016). Further research will be needed to discover the mechanisms underlying language transfers in early childhood, in order to shed some light on how the brain organises information sharing when two systems compete for resources.

Study 2: Predicting Vocabulary Scores in Bilingual Toddlers

The second key aim of the current paper was to develop a model of English and Additional Language vocabulary to be used with UK-raised bilinguals learning one of the target Additional Languages. To this aim we investigated which predictors should be included in this model, out of an extensive inventory which we summarise below (see also Table 1).

Relative amount of exposure to languages. As expected, the most robust predictor of English and Additional Language vocabulary was found to be the relative amount of exposure to child-directed English versus the Additional Language (e.g., Hoff, 2003). This was predictive of both comprehension and production, with greater exposure to English increasing vocabularies in English, and reducing them in the Additional Language. Similarly, the proportion of English/Additional Language spoken *between* parents was also strongly predictive of comprehension and production in English, in line with previous studies showing that children encode information from overheard speech (Akhtar, 2005; Shneidman, Arroyo, Levine, & Goldin-Meadow, 2013).

Mode of Exposure. We explored the predictive value of factors related to the source, properties, and status of the mode of exposure to English and the Additional Language. Only the two predictors related to the properties of the input survived our analytical process - yet statistical criteria prevented their inclusion in the final predictive models: the proportion of native English spoken to the child by parents, and the degree of language use consistency in parents' speech.

Source of each language. First we examined whether the fact that two parents were native Additional Language speakers, or only one, had an impact on vocabulary knowledge, as

was suggested by findings by Barrena et al. (2008). This factor did not have any significant impact in our data, perhaps because much of the variance associated to this variable was already apportioned to the relative amount of exposure to each language.

Second we looked at whether the total number of native English speakers, or the total number of Additional Language speakers around the child would affect word knowledge in each language. This was previously found by Gollan et al (2015), testing 8-year-old Hebrew-English bilinguals in a picture naming task, who reported that children interacting with more Hebrew speakers could name more pictures in this language (see also Place & Hoff, 2011). However Place and Hoff (2016) only reported modest effects of the number of speakers in Spanish-English 30-month-olds, and only in Spanish. Here we did not find any significant advantage for being surrounded by many native speakers. Further research is needed to clarify the impact of this factor on vocabulary growth, as it is of major theoretical interest to delineate the role of input variability in phonological development (Rost & McMurray, 2009).

Thirdly we found no effect of the time spent in daycare (in each language) on vocabulary development. Given that daycare attendance in monolingual children tends to benefit only those from low SES (Côté et al., 2013), and given that our sampling failed to capture a significant portion of low SES families, it is perhaps not surprising that we did not report any effect of this factor.

Finally, regarding the effect of the number of older siblings, predictions from the literature were mixed, with monolingual data pointing to larger vocabularies overall in first-born children (Huttenlocher et al., 2010), and bilingual studies suggesting larger English vocabularies for North American bilingual toddlers with older siblings (Bridges & Hoff, 2014). Here we did not report any significant effect of this factor, which perhaps is related to the fact that in our data collection we did not distinguish between toddlers who had school-aged siblings and toddlers whose siblings were still too young. Indeed Bridges and Hoff (2014)

reported that it was school-aged siblings who were mainly responsible for an increase in the proportion of English spoken at home and the resulting increase in toddlers' English vocabulary.

Properties of the input. We found that English production increased with the proportion of native English parental input, in line with previous findings that native speakers provide more supportive linguistic input than non-natives (Hammer et al., 2012; Place & Hoff, 2011, 2016; although see Paradis, 2011, who found no effect). Fernald (2006) suggested that non-native speech might hinder the development of phonological categories, but this appears contradicted by the fact that we did not report any *negative* effect of non-native input on vocabulary. Our findings may be better accounted for by non-native speakers using less varied vocabulary than native speakers, as suggested by Hoff et al. (2013).

Finally, we found that the degree of language use consistency in parents' speech was positively associated with English production, so that children whose parents used a mix of the two languages knew more words in English, perhaps because more code-switching increases the relative frequency of the majority language. However, previous findings on the impact of within-speaker consistency (measured at the sentence level as in code-switching, or at the discourse level as in the degree of language use consistency) suggest mixed outcomes: Byers-Heinlein (2013) reported a detrimental effect of code-switching at 18 months, but only marginally at 24 months, whereas Place and Hoff (2016) reported no robust relations between the degree of language use consistency in mothers' speech and language development at 30 months. In addition to an age-related explanation, we argued that the discrepancy between these two prior studies would be due to language distance, as Byers-Heinlein tested mainly distant languages learners whilst Place and Hoff examined close languages learners. To examine whether the positive effect found in the current study was different for close and distant language learners, we ran separate post-hoc regression analyses on the median split of children

learning close (N = 240) and distant (N = 132) languages (based on phonological overlap). For distant language learners, the degree of language use consistency in parental speech did not predict comprehension nor production, once corrected for exposure to English and the proportion of English in overheard speech. However, for children learning close languages, the degree of language use consistency in parental speech contributed significantly to improve English production scores (main model: F (4, 239) = 11.34, p < .001; standardised *ß* for language mixing = 0.15, p = .016; no effect for comprehension, or for Additional Language scores). These results do provide some preliminary support for our original hypothesis: distant language learners perform poorly with language mixing input measured through code-switching (Byers-Heinlein, 2013) or show no effect (this study), while close language learners either benefit from language mixing measured through language use consistency (this study) or show no effect (Place & Hoff, 2016). The explanation behind these findings would be that frequent switches from one set of representations to another in the case of distant languages impinge on cognitive resources, whereas close languages activate overlapping representations and therefore do not necessitate an increased cognitive flexibility.

Status of the Additional Language (societal vs. minority). We hypothesized that the societal status of the Additional Language would have consequences for children's achievements in the two languages, because of people's more positive attitude towards bilingualism, and also because of a more balanced exposure to the two languages in everyday life. However, Welsh-English toddlers in Wales did not have significantly higher vocabulary scores in their two languages than the other bilinguals, whether these scores were corrected by situational factors or not. In the literature, Welsh-English bilinguals' vocabulary in English tends to be lower than English monolinguals' (Rhys & Thomas, 2013) at the ages of 7 and 11 years, with a strong effect of language dominance (as defined by exposure and family language) and a complex relation between home language exposure, SES, and age (Gathercole, et al.,

2016). Our study shows that this "bilingual difference", previously found in Welsh, is similar to that of any other bilingual minority group in the UK, at least at the age of 24 months. This point contributes to strengthen the generality of the UKBTAT model for predicting language outcomes in any incoming bilingual toddler growing in the UK. Importantly, the fact that there was no effect for the societal status of the Additional Language being learned implies that our findings apply both to bilingual toddlers learning a minority language and also to those learning two languages which have more equivalent societal status, such as perhaps English and French in Canada.

Demographic factors (gender and SES). As expected, gender was found to be a reliable predictor of production vocabulary in both languages, with girls producing more words than boys (Huttenlocher et al., 1991); no effect was found on comprehension, which is also in line with previous findings (Eriksson et al., 2012).

Contrary to our initial expectations, SES effects were absent in our data, in keeping with prior studies which also had much reduced sampling at the low end of the SES spectrum (Fenson et al., 2007; Hamilton et al., 2000). The effects of these missing low SES children on the accuracy of UKBTAT norms are unclear. If they had been fully represented then we might have expected reduced vocabulary scores for low SES children, meaning that the current UKBTAT model would over-refer low SES children, and under-refer high SES children. There are a number of dangers with this assumption however, beyond those inherent in assuming that SES would have reached significance if provided with a fully balanced spectrum of samples. Fernald et al. (2013) reported that their SES effect, one of the strongest found in the literature, was stronger at the lower end of the SES spectrum. We confirmed this with a re-analysis of data points retrieved from Fernald et al's published scatterplot using the application Plot Digitizer (www.plotdigitizer.sourceforge.net), first reproducing the originally reported correlation of .34. We then went on to split this data at median SES and analyse high and low

SES data separately, finding that the correlation of SES with CDI comprehension was significant in low SES children (r = .42, p = .001), but not in high SES children (r = .27, p > .05). The nonlinearity of this effect means that it would be poorly modelled by linear equations, such as those used in the UKBTAT.

In summary, the three factors that were identified to be the key predictors of vocabulary knowledge in 2-year-old bilinguals learning English and one target Additional Languages were the relative amount of exposure to English versus the Additional Language in child-directed input (English and Additional Language production and comprehension) and parental overheard speech (English production and comprehension), and gender (Additional Language and English production). The resulting models were built into the UKBTAT assessment tool, which fulfilled the objectives of Study 2. The UKBTAT uses measures of English vocabulary on the 100-word Oxford Short Form CDI, and if the child learns one of the target Additional Languages, the full Additional Language CDI is administered, from which the 30 words common to all CDIs are extracted to derive predictions. In addition, the features of the UKBTAT (online, based on parental report and interview) are particularly appealing for an initial screening at such an early age, where language assessments are notoriously challenging and yet very desirable to plan appropriate intervention.

Application of the UKBTAT with Bilingual Children with Non-Target Additional Languages

We created the UKBTAT norms for assessing the vocabulary of any bilingual 24-month-old learning British English and an Additional Language, regardless of which Additional Language it would be. These norms are available in British English, and in the Additional Language when this is part of our target languages. The recommendations of professional bodies are that bilinguals should be assessed in both their languages as there is a, quite justified, risk that an individual child's situational factors may render an assessment

performed on only one language unrepresentative of their overall linguistic capability. Unfortunately, for pragmatic reasons due to the heterogeneity of additional languages spoken in the world - and in the UK in particular - this recommendation can only be followed in a relatively small minority of cases, with bilinguals generally only assessed in the majority language. Therefore, in this project we have sought to identify the situational factors that affect bilingual acquisition and quantify their contribution to improve the quality of English only testing, in the case of UK-raised toddlers. We have shown that only the relative amount of English exposure in child-directed input and overheard speech interacted with test language in the determination of vocabulary scores, thus we have no evidence that situational factors beyond these apply differently to the two test languages. This means there is no reason to suppose that Additional Language scores provide useful information beyond that of an English score, once corrected for these factors. Therefore, for clinical purposes, information provided by English vocabulary scores alone appears to provide an effective proxy for overall linguistic attainment, and can be safely used as normative data for any new incoming bilingual toddler. It is our hope that this pragmatic approach to bilingual screening will be generalised to other countries who are, like the UK, facing a growing number of bilingual infants from an increasingly heterogeneous background of languages.

Limitations and Future Directions

An important limitation of this study pertains to the relatively small proportion of children with a low level of exposure to English, either because of SES characteristics, date of arrival in the country or simply family characteristics. Regarding the poor representation of low SES families in our sample (only 4 children out of the cohort scored on the lowest income band): as noted earlier, the current UKBTAT model would probably over-refer low SES children and, possibly, under-refer high SES children; only further research will allow to address this issue satisfyingly. Regarding date of arrival in the country, it must be noted that

we did not include age of acquisition in the analyses, because the vast majority of toddlers were born in the UK (see participants section), which we took as a reasonable proxy for simultaneous acquisition, but it would be unreasonable to use the UKBTAT for a child whose age of acquisition of British English is only a few months prior assessment: not only is the effect of age of acquisition a very established predictor of bilingual development (Flege, Yeni-Komshian, Liu, 1999), but the Plymouth Language Exposure Questionnaire, which is implemented in the UKBTAT, does not provision (yet) for variation in the age of acquisition. Finally, the characteristics of the families included in the survey were such that only 7% of the 430 children had less than 10% exposure to English (as measured by the Plymouth Language Exposure Questionnaire). This probably relates again to the poor sampling of low SES families, and calls for caution in the use of the UKBTAT when assessing toddlers with very low English exposure.

In the same vein, another limitation relates to the mode of calculation of exposure to each language, which was here exclusively focused on the *relative* exposure (as most exposure questionnaires do: e.g., De Houwer & Bornstein, 2003; Paradis, 2011), and not on the *quantity* of exposure. That is, a child who scored 60% of English on our measure of exposure could experience a few hours of cumulated English speech per week or a few dozen, depending on the communication style of her caregivers. Measures of quantity (obtained through naturalistic recordings) are known to predict language growth in young monolinguals (Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010), and thus a full measure of exposure should encompass both relative and absolute measures of language input. Additional research will be needed to evaluate if quantity of bilingual exposure explains an additional portion of the variance in explaining vocabulary growth.

Another limiting factor relates to the age of toddlers, which we selected to be two years old, as an important milestone for parents, researchers and practitioners, in time for preschool

referral if needed. Whether our key findings would hold for younger and older children remain to an empirical question, but we anticipate that measures of exposure and gender would remain central predictors of bilingual development throughout development (see Hyde & Linn, 1988 for effects of gender through development). What is more interesting is how linguistic distance metrics would predict language development before and after the age of two. Generally speaking, it is likely that the type of distance measures that would modulate bilingual language growth would depend both on the age of the child and to-be-assessed language skills, with perhaps phonological overlap metrics becoming less influential as grammar and morphology grow in the child. We are currently exploring these claims as a substantial subset of the cohort tested in this paper were re-assessed at age 3 with the same tools.

Finally, the language-specificity of our findings implies an obvious caveat, as the UKBTAT is currently valid only for UK-raised toddlers who learn British English as one of their languages. However there is no reason to think that the key results – that linguistic distance is predictive of vocabulary growth, and that a useful model of the early lexicon needs only a handful of predictors – would be language-specific. Of particular interest would be to replicate this cohort study in North America, as minimal adjustments to the 100-word Oxford Short Form CDI would be required, allowing the problem of assessing a growing number or bilinguals from heterogeneous populations (Statistics Canada, 2011; United States Census 2020) to be addressed.

Conclusion

The current research included three key strands of research questions. In the first study we established that linguistic distance between the two languages plays a role in predicting bilingual toddlers' vocabulary at age 2, a result which moved the field forward. Given the interweaving of word-learning abilities and cognitive processes in development (Baddeley, Gathercole, Papagno, 1998), the impact of linguistic distance might extend beyond language

acquisition, and could constitute a key factor in the study of the debated bilingual cognitive advantage (Bialystock, 2009; Paap & Greenberg, 2013).

In Study 2, which had an applied orientation, we identified the amount of exposure to each language in child-directed and overheard speech as key predictors of vocabulary development in bilingual toddlers, along with infant gender, and developed a model of the bilingual lexicon usable for assessment purposes, the UKBTAT. Finally, in Study 3, we explored the feasibility of assessing a bilingual toddler's vocabulary in the majority language only, in order to provide professionals working with young children with a practical solution when encountering bilingual children from heterogeneous backgrounds, as is the case is most of the world. The validity of this approach was successfully demonstrated when the UKBTAT model was applied to a new cohort of bilinguals learning a different Additional Language to our original target languages. It is our hope that the UKBTAT will enhance the early detection of children at risk of language delays in the growing UK bilingual population, and perhaps in other countries where the same approach could be applied.