How Bilingual Language Experience Shapes the Infant Lexicon: Implications for Word Learning and Lexical Access

Serene Siow

Lady Margaret Hall

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2022

7th October 2022

Contents

How Bilingual Language Experience Shapes the Infant Lexicon: Implications for Word Learning and Lexical Access

Serene Siow

Abstract

In this thesis, I investigate how the words known by bilingual toddlers affect their learning of new words. I explore phonological and semantic connections both within and between languages. I analyse vocabulary questionnaire data collected via parent report for bilingual and monolingual toddlers growing up in the UK. Firstly, I show that while bilinguals' English vocabulary size lags behind that of monolinguals, conceptual vocabulary size is similar across groups. Notably, when words across both languages are added together, bilinguals understand many more words compared to monolinguals. Bilinguals' vocabularies include words with the same meaning across languages (translation equivalents). I found that if a word is known in one language (e.g., English "dog"), they have a high likelihood of also learning its translation equivalent (i.e., Spanish "perro"). Additionally, the acquisition of translation equivalents may be supported by systematicity between phonology and meaning. Even monolingual adults are able to make use of cross-linguistic similarity to guess the meaning of unfamiliar foreign words. I show that bilingual toddlers learn cognates (translation equivalents with phonological overlap) earlier than non-cognates, suggesting that toddlers can make use of cross-linguistic similarity to support word learning. Toddlers also display sensitivity to semantic connections in their language(s). Using semantic network analyses, I show that high frequency words and words that have a large number of associative connections to other words are learnt earlier by toddlers than words with low frequency and few connections to other words. The words within a toddlers' lexicon also have dense connectivity with each other, much more than would be expected if associative patterns in the learning environment had no impact on word acquisition order. Findings from this thesis highlight how bilingual toddlers can capitalise on phonological and semantic connections between words to facilitate learning two languages in parallel.

Chapter 1

Introduction

Word learning is a difficult process that involves extracting the spoken word form from the continuous speech signal, inferring the word's meaning, and storing the lexical representation for later retrieval of the word form and meaning. Studies on lexical retrieval have supported the idea that words are stored in the lexicon with links to phonologically-related words (e.g., Marslen-Wilson & Welsh, 1978; Slowiaczek & Hamburger, 1992; Meyer & Damian, 2007) and semantically-related words (e.g., Moores, Laiti, & Chelazzi, 2003; Meyer, Belke, Telling, & Humphreys, 2007). Word learning is even more complicated for bilinguals, who need to learn two sets of words, one in each language. Luckily, languages overlap in some elements which may help bilinguals learn words more easily. Languages share words that have the same meaning (also known as *translation equivalents*), and also typically share part of their phonological inventories. Bilinguals subsequently may form connections between words not only within a language but also between their two languages to support word learning and word retrieval. The degree of interconnectivity between the two languages in bilinguals' word learning and lexical organisation continues to be a topic of interest in bilingual research.

In this thesis, I explore how the words that bilingual toddlers already know can affect their learning of new words. The thesis can be split into three broad sections – the first evaluating bilinguals' vocabulary growth, the second studying the effect of phonological connections on word learning trajectories, and the third studying the effect of semantic connections on word learning trajectories.

1.1 Lexical activation

When studying auditory recognition of words, it is important to consider what happens during online processing by active listeners of speech. Spoken word processing is by nature incremental due to the temporal speech signal, meaning that the sounds in a word are presented sequentially to the listener (and therefore also perceived sequentially). Mispronunciations occurring in the second syllable are detected faster than mispronunciations in the first syllable (Cole & Jakimik, 1980). Cole and Jakimik argued that this is due to the first syllable having a stronger effect on lexical retrieval. When the first syllable is correctly presented, it is easier to retrieve the correct lexical representation (and therefore be able to compare the mispronunciation against the correct pronunciation). Theories of lexical activation involving the simultaneous activation of multiple possible candidate words have been proposed to explain the process of spoken word recognition (Marslen-

Wilson & Welsh, 1978) and speech production (Levelt et al., 1991). Under these models, the better a candidate matches the auditory input, the higher the activation it receives. The candidate with the highest activation will be selected as the best fit for the auditory input. Research have also supporting the idea that the meanings of the candidate words will also activated (Marslen-Wilson, 1987) and multiple meanings may be activated simultaneously in the case of semantically-ambiguous words (Swinney, 1979).

There is evidence suggesting that when both adults and toddlers see a familiar object, they automatically and rapidly retrieve its name and semantic information. This process is known as implicit naming. Meyer et al. (2007) utilised a visual world paradigm to study implicit naming in adults. After seeing a picture presented in silence, participants were asked to identify whether the previously-shown object was present in a subsequently-presented 4-picture array. They studied participants' responses on target-absent trials where a distractor was a homophone of the target. Despite the distractors not being named aloud, participants were slower to reject target-absent trials when a homophone distractor was present in the array than when all objects were unrelated to the target. Homophone distractors were also fixated on longer than unrelated distractors. This suggests that participants automatically activated the names of the target item and the items in the array, and the overlapping phonological information between target and homophone distractor subsequently interfered with target identification. In another study, Meyer and Damian (2007) found that presenting phonologically-related target and distractor picture pairs together resulted in faster target picture naming in adult participants. This was found for homophones, pairs that shared the same two onset phonemes, and pairs that shared the rhyme. They proposed that when viewing a target-distractor pair, participants implicitly name both items, and the activation of shared phonological information in the phonologically-related condition allowed the target to be activated and produced faster. There has also been evidence of implicit naming in toddlers. A study by Mani and Plunkett (2011) used a phonological priming paradigm where toddlers were first shown a silent prime image, followed by a target word that was either phonologically-related or unrelated to the prime. They found that the prime's name (e.g. "flower") influenced the recognition speed of phonologically-related targets (e.g. "fork"), even though the prime's name was never presented aloud. This suggests that toddlers' implicitly named the prime after seeing the picture. This was done without explicit instructions to name the prime.

Automatic naming has been studied not just with monolinguals, but with bilinguals as well. A unique element of bilinguals' language experience is that each object in their environment has two names, one in each language. Instead of simply learning that a dog is known as a "dog", an English-Spanish bilingual learns that the animal dog is called a "dog" and also a "perro". These dual-language word pairs which share a referent are referred to as translation equivalents (TEs). The non-selective bilingual lexical access hypothesis posits that bilinguals implicitly name objects in both languages simultaneously, even when task demands only require one language (Costa, Caramazza, & Sebastian-Galles, 2000). Costa et al. found that cognates were named faster than non-cognates by bilingual adults, suggesting facilitation from the item's name in the non-test language. The Bilingual Language Interaction Network for Comprehension of Speech (BLINCS) model (Shook & Marian, 2013) posits that cognates lie in the intersection between two languages due to their shared meaning and phonological overlap. The overlapping phonological elements of cognate pairs feed activation to each other, facilitating faster production of cognate words (Costa et al., 2000).

Non-selective lexical access is not restricted to cognates. In a picture-word matching task, bilinguals showed a reduced N400 as measured by event-related potentials (ERPs) for mismatched trials where the picture (e.g., of a beach) was paired with a word in the test language that sounded similar to the picture's label in the non-test language (e.g., presented word "plaid" sounds similar to the French word "plage" which means beach), as compared to mismatched trials where the picture and word were unrelated (Desroches, Friesen, Teles, Korade, & Forest, 2022). This difference was not observed in monolinguals. Desroches et al. took this as support for non-selective access of both languages by bilinguals, even when only one language was required by the task. Spivey and Marian (1999) studied the effect of translation distractors, where the distractor's name in the non-test language had phonological overlap with the spoken target word. Bilingual adults looked more at translation distractors than unrelated distractors, despite the overlap being in the non-test language. This supports the idea that bilinguals activate both languages even under circumstances that only require one language to be used. Non-selective lexical access has also been found in young language learners (21 to 43 months old). This is supported by findings of interference on target recognition in a cross-linguistic translation priming paradigm, where the target word in toddlers' L1 was phonologically-related to the prime's name in their L2 (Von Holzen & Mani, 2012). As priming effects would not be able to occur without lexical activation of the translation (which was not spoken aloud), the observed interference effect was taken as support for non-selective lexical activation of both languages in bilingual toddlers.

In the initial draft of this thesis, an in-lab eye-tracking experiment investigating the non-selective lexical access hypothesis in bilingual toddlers was planned. I aimed to investigate the simultaneous implicit naming of an objects' names across both languages even when the object is presented visually without accompanying speech. This would have been tested via a phonological priming paradigm. It has previously been found that a silent prime image can cause lexical interference for the recognition of a phonologically-related target (operationalised by lower proportion of target looking) when activated competitors apply inhibition on the target (Mani & Plunkett, 2011). The non-selective lexical access hypothesis posits that the shared phonological elements in a cognate pair feed activation to each other, strengthening the activation of the word (Costa et al., 2000). Cognates are therefore expected to be stronger primes than non-cognates. If target looking times are lower following cognate primes than non-cognate primes, it would give support that the non-selective lexical access hypothesis is also applicable to bilingual toddlers. Unfortunately, the restrictions associated with the COVID-19 pandemic made it difficult to recruit participants for in-lab studies. This study had to be excluded from the final version of this thesis.

1.1.1 Matching phonology to lexical representation

When comprehending and producing language, we need to be able to accurately retrieve the meaning of words. This involves matching each word's phonological form to the correct lexical representation in the mental lexicon and also differentiating the word from other words that may sound similar. Listeners need to match the phonological form [kæp] to the lexical representation of "cap" while being clear that it does not correspond to "cup", "cat" or any other similar-sounding words. Listeners' ability to learn and differentiate phonological neighbours (words that differ on only 1 phoneme) has been of great interest in the literature of spoken word recognition. Infants at 6 months old are sensitive to a wide range of sound contrasts, including contrasts that are not differentiated in their native language (Werker & Tees, 1984). Sensitivity to non-native contrasts subsequently declines in the first year of life. This perceptual reorganisation of phoneme categories has been proposed to facilitate more efficient recognition of contrasts relevant to the learners' native language.

Various studies have supported the idea that toddlers have high sensitivity to the phonetic features in the lexical representation of familiar words, similar to adults. Fourteen-month-old infants were able to discriminate between familiar words "ball" and "doll" which differed only in the onset phoneme minimal pair [b] and [d] (Fennell & Werker, 2003). Swingley and Aslin (2002) showed that infants are sensitive to mispronunciations of familiar words. While 14-month-old infants showed above chance target looking in response to both correct pronunciations and mispronunciations of the target word, they found it more difficult when the target word was mispronounced. Toddlers (15, 18 and 24 months old) were sensitive to both onset consonant and median vowel mispronunciations in familiar words with CVC structure (Mani & Plunkett, 2007). Even 12-month-old infants show sensitivity to both consonant and vowel mispronunciations (Mani & Plunkett, 2010b). In both of these studies by Mani and Plunkett, there was no significant difference in the degree of sensitivity between consonant or vowel mispronunciations. However, an advantage for consonants during word recognition has been found in several other languages, including French (Nishibayashi & Nazzi, 2016; Von Holzen, Nishibayashi, & Nazzi, 2018) and Italian (Hochmann, Benavides-Varela, Nespor, & Mehler, 2011). In contrast, there has been findings of a vowel advantage for languages such as Danish which have high numbers of vowels relative to consonants (Højen & Nazzi, 2016). The differing pattern across different languages, along with studies showing vowel advantage in younger French and Italian infants (Benavides-Varela, Hochmann, Macagno, Nespor, & Mehler, 2012; Hochmann, Benavides-Varela, Fló, Nespor, & Mehler, 2018; Nishibayashi & Nazzi, 2016), lend support to the theory that consonant and vowel biases may form during development in response to the language the infant is exposed to. These studies highlights how potential differences between languages need to be taken into consideration when conducting cross-linguistic research. Another factor that infants' mispronunciations have been found to be affected by is the distance between phonemes White and Morgan (2008). When tested with mispronounced familiar words, 19-month-old toddlers showed graded sensitivity to mispronunciations that differed on 1 phoneme feature, 2 phoneme features or 3 phoneme features. They also showed similar levels of sensitivity to 1-feature changes of place, manner or voicing. Sensitivity was also similar between mispronunciations with different combinations of 2 feature changes. This graded sensitivity to the number of phoneme feature changes suggests that toddlers' lexical representations of familiar words are mature and that toddlers are sensitive to sub-segmental detail in lexical processing.

However, it is not always optimal to summarily reject all inaccurate phonetic representations in a word. Noisy listening conditions, accented speech and even natural variation between speakers are common occurrences in the real world that can result in spoken input that does not perfectly match existing lexical representations in the listener's lexicon. To successfully understand speech under these challenging circumstances, listeners need some degree of flexibility in their lexical inference. This could be guided by exposure to variability in the input. If a listener learns that a specific phoneme variation is recurring and characteristic of a given speaker (as frequently occurs in the context of accented speech), they may start accepting those variants as members of the appropriate phoneme category to aid word recognition (Kraljic, Samuel, & Brennan, 2008; Kraljic, Brennan, & Samuel, 2008). van Heugten, Paquette-Smith, Krieger, and Johnson (2018) found that 18-month-old English-learning toddlers recognised known words in an unfamiliar foreign accent (French-accented English) but not when the known words were mispronounced, suggesting that there is a specificity to the acceptance of phoneme variation even among toddlers.

1.2 Phonological connections in the lexicon

Connine (1994) used the terms horizontal similarity and vertical similarity when discussing two similarity-driven processes during spoken word recognition. Vertical similarity was used when referring to the degree of similarity between the input word and possible candidate words, inclusive of cases where the input does not perfectly match up to any lexical items. This effect is studied by observing participant responses when identifying mispronunciations and listening under noisy environments (Cole, Jakimik, & Cooper, 1978; Marslen-Wilson & Welsh, 1978). When hearing a perceptually-ambiguous input word, listeners activate multiple possible candidate words that have the closest similarity to the input (Connine, Blasko, & Wang, 1994). Horizontal similarity was used when referring to phonological similarity between lexical items. When hearing an input word, other candidate words that are closely similar to the input (i.e. neighbours) would also be activated as a result of horizontal similarity. This was linked to phonological neighbourhood effects on lexical retrieval.

The role of phonological competition in lexical retrieval forms a core assumption of several computational models of spoken word recognition, including the Cohort Model (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987), TRACE (McClelland & Elman, 1986) and Shortlist (Norris, 1994). The presentation of a word's first phoneme activates all relevant words that could be the input word. As the number of phonemes increases, the number of competitors decreases, until it reaches the target word's uniqueness point. A word's uniqueness point refers to the point in the word's incremental presentation where it becomes unambiguous which candidate word should be selected. For example, the phoneme cluster [kæptɪ] could be the start of "captain" or "captive", but [kæptɪn] would be unambiguously "captain". In this scenario, the word-final [n] would be the uniqueness point. Some longer words may have a uniqueness point prior to the word-final phoneme, but for many words, the uniqueness point would be at the end of the word (Luce, 1986). In the updated Cohort Model's activation-based candidate selection (Marslen-Wilson, 1987), the candidate word with the highest final activation at the end of the word presentation phase would be selected. Importantly, other mismatched candidate words are not dropped fully from the cohort, but instead are simply less strongly activated than the selected candidate word. Simultaneous activation of candidate words has been used to explain semantic priming effects where the words "captain" and "captive" are equally strong primes when the target probe is presented prior to these words' uniqueness point (Marslen-Wilson, 1987). Candidate words gain activation when phonemes match up to the input, and lose activation from mismatching phonemes. Spinelli, Segui, and Radeau (2001) found facilitation effects in a phonological priming paradigm when the monosyllabic prime overlapped fully with the bisyllabic target's initial syllable (e.g., "ver" - "vertige"), but not when the prime was also bisyllabic and overlapped with the bisyllable target in only one of the two syllables (e.g., "verger" - "vertige"). This finding was notable as the number of overlapping phonemes between primes and targets was the same for the two conditions, but the number of mismatching phonemes differed. This suggests that the mismatching phonemes in "verger" and "vertige" suppressed the activation of "vertige" as a candidate word, resulting in slower lexical retrieval of "vertige".

The co-activation of candidate words is supported by studies of lexical access using phonological priming. Both adults and toddlers are proposed to group words according to shared phonology and shared semantics, which facilitates the retrieval of words during word recognition and production. Phonological similarity between the target and prime or between the target and distractors in the task affect target identification times in both adults (Slowiaczek & Hamburger, 1992; Meyer & Damian, 2007) and toddlers (Mani & Plunkett, 2011). This suggests that in addition to activating the name of the presented object, activation is also given to other phonologically-related words. For example, when they see a “flower”, participants automatically activate words which share an onset phoneme, like “frog”, “fish” and “fork”. Interestingly, the effect of phonological similarity seems to differ in adults depending on the degree of similarity. Slowiaczek and Hamburger (1992) differentiated between facilitation effects found for prime-target pairs that share one onset phoneme, and inhibitory effects found for prime-target pairs that share three initial phonemes. They proposed that facilitation effects stem from prelexical phonological activation from the shared onset phoneme. In contrast, high overlap of three overlapping phonemes prompts lexical activation of matching competitors, resulting in inhibitory effects from lexical competition. Dufour and Peereman (2003b) delved deeper into this pattern of facilitation and inhibition, showing that inhibitory priming effects can be found both for words with 3-phoneme overlap and 2-phoneme overlap, but only when there is only 1 mismatching phoneme. Trials with 2-phoneme overlap did not show inhibitory effects if there were also 2 mismatching phonemes. Dufour and Peereman suggested that inhibitory effects from lexical interference are dependent on a higher ratio of matched phonemes to mismatched phonemes, which prompts stronger activation of relevant competitors with high overlap. This theory is also compatible with the findings from Spinelli et al. (2001) where neither facilitation nor inhibition was found for bisyllabic prime-target pairs that only overlap on one of the two syllables. Taken together, these studies show that lexical interference during word recognition occurs in scenarios where competitors are strongly activated via high phonological overlap with the target word, resulting in slower target recognition.

In contrast to studies with adults, lexical interference effects have been found for toddlers even for prime-target pairs that share only the onset phoneme. This has been proposed to be linked to toddlers' immature inhibitory control, where a poor ability to inhibit any competitors results in slower target recognition speed. Under the lexical interference hypothesis put forward by Mani and Plunkett (2011) for toddlers, the presentation of a prime image activates the name of the prime (e.g., "flower") and other words that share its onset phoneme (e.g., "frog", "fish", "fork"). Subsequently, the successful retrieval and recognition of the target word (e.g., "frog") requires those other competing activated words to be inhibited. When prime and target are phonologically related, the phonological information from the spoken target word will add more activation to the same set of competitors. The activation of competitors is therefore stronger when prime and target are phonologically related, requiring more inhibition to overcome. There is also evidence that the lexical interference effect develops as the lexicon grows larger in young language learners. Mani and Plunkett (2010a, 2011) found that 18-month-old monolinguals toddlers did not show lexical interference during target recognition following phonological priming, while 24-month-old toddlers did. They linked this to differences in vocabulary size, with a sufficiently large vocabulary size being needed to trigger lexical competition. When splitting participants by vocabulary size, 18-month-old toddlers with larger vocabulary sizes showed lexical interference effects but those with smaller ones did not. Mani and Plunkett proposed that the lexical interference effect exerted by competitors increases with the number of competitors activated, supported by findings that lexical interference effects were stronger in trials where the target had larger cohort size.

In bilingual toddlers, Von Holzen and Mani (2012) showed that phonological primes (where the spoken prime word had phonological overlap with the target word) resulted in facilitation during target identification, while translation primes (where the *translation* of the spoken prime word had phonological overlap with the target word) resulted in interference. They proposed that the facilitation observed in the phonological prime condition could be attributed to pre-lexical priming effects via shared phonological information between the spoken prime word and spoken target word. In contrast, they attributed the interference effect from translation primes to lexical competition. This interpretation is consistent with that of previous studies with phonological priming such as Slowiaczek and Hamburger (1992) and Mani and Plunkett (2011).

The previously-mentioned phonological priming study that was meant to be included in this thesis would also have probed the effect of bilingual phonological neighbourhoods. Under the non-selective lexical access hypothesis, cognates are posited to not only activate the two phonologically-overlapping translation equivalents, but also other phonologically-related neighbours. Due to the phonological overlap between cognates, the neighbours activated across both languages would also have phonological overlap with each other. This large combined phonological neighbourhood is posited to exert additional lexical interference on target recognition. Importantly, the effect of phonological neighbours from the non-test language should only apply to bilingual toddlers (who are theorised to activate words that they know in both test and non-test languages) and not to monolingual toddlers (who only know words in the test language). Therefore, if a difference between cognate and non-cognate trials is found between bilinguals but not monolinguals, it would support the theory that bilinguals' two languages are simultaneously activated and influence word recognition through shared phonological connections.

1.2.1 Quantifying phonological similarity

Given extensive research on both within-language and cross-linguistic phonological similarity, it is useful for researchers to have a reliable measure for comparing the degree of similarity between words. Similarity between word pairs varies on a spectrum and can differ depending on the medium used. Historical perspectives on cross-linguistic similarity often focus on common etymology, while behavioural studies may use orthographic or phonological similarity depending on the task demands. Calculations that produce a quantitative score of overlap or distance between two words are widely used. The most prevalent method is Levenshtein distance (Levenshtein, 1966), which counts the number of additions, subtractions and substitutions needed to change one string to another. A larger value indicates that more edits are needed. It can be applied to words' orthographic form (spelling) or phonological form (transcriptions of pronunciation). Under this method, cognates are defined as semantically-equivalent word pairs with low Levenshtein distance score. When comparing between different word pairs, it is common to utilise a standardised score which controls for word length. A numeric cut-off is commonly applied for cognate classification, though some research may also use Levenshtein distance as a continuous measure.

With infant research specifically, orthographical similarity is irrelevant because infants have not yet learnt to read – infants can only recognise cognates through phonological similarity. While orthography and phonology tend to be correlated in many languages, the overlap is weaker in languages with deep orthography such as English than it would be for shallow languages like German (Schepens, Dijkstra, Grootjen, & Van Heuven, 2013). It is therefore more appropriate to utilise a measure of phonological similarity for infant research instead of relying on orthographic similarity. Floccia et al. (2018) calculated the Levenshtein distance between the phonological transcriptions of cross-linguistic translation equivalents. They estimated the overall phonological similarity between each of their tested language pairs using the average Levenshtein distance of translation equivalents in the CDI.

While Levenshtein distance can provide a general estimate of phonological similarity, there are unique challenges in measuring phonological similarity. Unlike differences in orthography that are discrete (a ‘k’ is not a ‘g’), sounds can vary in their similarity ([k] sounds closer to [g] than [u]). An attempt to address this issue is proposed in algorithms like Kondrak’s (2003) ALINE and P. Li and MacWhinney’s (2002) PatPho, but there is yet to be a single standardised method used by all researchers.

Given the availability of multiple options, I will evaluate the accuracy of two algorithms – Levenshtein distance and ALINE in Chapter 3. Both of these algorithms have the advantage of being re-implemented as R (R Core Team, 2013) packages, making them easy to use. The algorithm that performs best out of these options will be used for defining cognates in Chapter 5. I will additionally evaluate the use of a spoken word recognition computational model in Chapter 4 for quantifying the phonological similarity between words. Models of spoken word recognition have the advantage over standard algorithms of phonological similarity of integrating lexical competition between other similar-sounding words.

1.2.2 Cross-linguistic phonological connections

A bilingual's two languages are tightly interwoven, influencing and complementing each other. A child's language experience is not only shaped by their language environment, but also the properties of the language(s) they are learning. One such property would be phonology, as words within a language and across languages share phonological links. Similarity between a language pair, encompassing aspects like phonology, morphology and grammatical structure, can facilitate the process of learning two languages (Floccia et al., 2018). Systematic links between words in both phonology and semantics can help the learning of new words and subsequent retrieval. In their review, Van Assche, Brysbaert, and Duyck (2020) highlighted three types of phonologically-related cross-linguistic words pairs – cognates (similar or identical orthography and phonology, same meaning; English "fish" and German "Fisch"); interlingual homographs (same orthographic word form, different meaning; for example English “pan” and Spanish “pan”), interlingual homophones (same pronunciation, different meaning; English “floor” and Spanish “flor”). Interlingual homographs and homophones are also commonly referred to as false cognates. Shook and Marian (2013) formulated a computational model that was able to extract the phonological properties that help differentiate two languages. The Bilingual Language Interaction Network for Comprehension of Speech (BLINCS) model was able to automatically separate English and Spanish words by their phonotactic properties to form distinct islands. Interestingly, cognates and false cognates clustered at the intersection between the two languages as a result of their phonological overlap.

Cognates have been found to exert a facilitatory effect on word recognition and word learning by bilinguals. Bilingual adults are faster at picture naming when the target name is a cognate across their two languages (Costa et al., 2000). This cognate facilitation effect can operate via phonological similarity even when the language pair has different orthographic script (Hoshino & Kroll, 2008). Cognate effects on word recognition have been found even in toddlers (Von Holzen, Fennell, & Mani, 2019). German-English bilingual toddlers were better at recognising L2 (English) words with higher phonological overlap with its L1 (German) translation equivalent. Interestingly, the effect of L1 facilitation on L2 word retrieval was stronger than L2 facilitation on L1 word retrieval. This asymmetry has been observed in various other studies showing that lower proficiency speakers benefit more from cognates than fluent speakers (Kroll, Michael, Tokowicz, & Dufour, 2002; Van Hell & Dijkstra, 2002; Pérez, Peña, & Bedore, 2010; Bultena, Dijkstra, & Van Hell, 2014; Allen, 2019). The asymmetry between L1 and L2 lexical access has been modelled in the Revised Hierarchical Model (RHM) (Kroll & Stewart, 1994), which focused on the bilinguals' production and recognition of translations. In an experiment reported in Kroll and Stewart (1994), they found that participants were slower to translate words from their L1 to L2 than when they were translating from L2 to L1. Kroll and Stewart suggested that lexical access of an L2 word, at least for less proficient learners and less frequent words, may be mediated by the L1 translation. Kroll et al. (2002) suggested that less fluent learners draw upon similarities between the L1 and L2 to help process L2 words. Even for fluent speakers, the simultaneous activation of both translation equivalents in a cognate pair (according to non-selective lexical access models) is posited to increase the activation of the target word through their shared phonology, facilitating lexical retrieval of cognates (Costa et al., 2000; Shook & Marian, 2013). An alternative hypothesis for the cognate facilitation effect, put forward by Sherkina-Lieber (2004), proposed that cognate facilitation may stem from a frequency effect, where the words are more easily retrieved because of their higher frequency of occurring in speech. Particularly, Sherkina-Lieber highlighted the effect of perceived frequency. They found that Russian-English bilinguals rate cognates as relatively more frequent than monolinguals do, while non-cognates' frequency ratings were not significantly different across groups. The phonological overlap between cognates may elicit the word in the other language, generating an illusion of higher frequency, which in turn facilitates word learning and retrieval. Regardless of the mechanism underlying it, findings of cognate facilitation suggests that cross-linguistic phonological similarity can help strengthen lexical representations that are still weak and aid lexical retrieval.

In Chapter 5 of this thesis, I will investigate the effect of cognates on the vocabulary trajectories of bilingual toddlers (12 to 36 months old). Findings from this study will support theories that toddlers are sensitive to the phonological similarities between their languages even from an early age and can use those similarities to learn words more easily.

1.3 Semantic connections in the lexicon

Words in a language also have links from their shared features, meaning and co-occurrences. When discussing these connections, the literature has generally revolved around taxonomic and associative connections. When two items belong to the same category, for example the food items “cake” and “bread”, they are considered to be taxonomically-related. We can operationalise taxonomic links between concepts using their shared defining features. A large database of shared features by McRae, Cree, Seidenberg, and McNorgan (2005), includes features corresponding to visual–form and surface (e.g., “has 4 legs”), visual-colour (e.g., “is blue), function (e.g., “used for cutting”), encyclopedic (e.g., “lives in woods”), tactile (e.g., “is sharp”) and taxonomic (e.g., “a mammal”). Semantic connections are established between objects that share one or more features. Meanwhile, some pairs are related by association, for example “juice” and “glass”, where the former is a drink and the latter is a container used to hold it and drink from. There are also many pairs that have both taxonomic and associative connections, for example “fork” and “spoon” which are both pieces of cutlery, have functional association (used together to eat) and are commonly seen together. Judgements of associative links have been derived from free-association tasks such as the University of South Florida Free Associations Norms (Nelson, McEvoy, & Schreiber, 2004) and the Small World of Words database (De Deyne, Navarro, Perfors, Brysbaert, & Storms, 2019). By collecting data across many participants about the first word(s) that come to mind in response to a given cue word, we can gain information about which words are associated with each other. The number of participants who provide a particular response can additionally be used as a proxy measure for the strength of association. Free association tasks also provide information about the forward and backward associations between concepts. Forward associations refer to responses that are elicited from a cue, while backwards associations refer to cues that elicit a given response. A word pair can also have both forward and backward associations concurrently. Both forward associations and backward associations have been shown to elicit priming effects, though the timings differ (Hutchison, Balota, Cortese, & Watson, 2008).

Lucas (2000) showed in a meta-analysis that both taxonomic primes and associative primes produced significant priming effects, but the effect size for associative priming (weighted average Cohen's d .47) is larger than that of taxonomic priming (weighted average Cohen's d .29). Additionally, priming was stronger for pairs with both associative and taxonomic links, over pairs with only taxonomic links. In a visual world paradigm where participants were asked to identify if a target (e.g., "key") was absent or present in a 4-picture display, distractors that were related by association to the target (e.g., "lock") were remembered better in a later recall task (Moores et al., 2003). Additionally, Moores et al. found that participants were less accurate and slower to reject target-absent trials when there was an associated distractor in the array. Meyer et al. (2007) found similar results. Meyer et al. tested the effect of phonological competitors and semantic competitors in the 4-picture array. When asked to identify if a target was present, both phonologically-related competitors (e.g., animal "bat" and baseball "bat") and semantically-related competitors (e.g. "shirt" and "trousers") resulted in slower participant responses. These independent effects of phonological connections and semantic connections highlight the structure of the mental lexicon. Words are linked by both phonological and semantic connections, which can help and hinder word retrieval.

A second in-lab experiment planned for this thesis would have probed the development of semantic connections in toddlers' lexicon using a backward semantic inhibition paradigm, replicating work by Chow, Aimola Davies, Fuentes, and Plunkett (2016) with a bilingual sample. This paradigm involves the initial activation of a target through semantic priming, followed by inhibition through an intervening category shift. If toddlers have formed a semantic connection between two concepts (e.g. train and helicopter), they should look more towards the train than an unrelated object (e.g. a flower) after seeing a helicopter as a prime. However, if an intervening object of a different category (e.g. cat) was introduced between helicopter and train, we would expect to observe less looking towards the train even compared to baseline looking preference. The effect of inhibition from an intervening category shift would suggest that toddlers have formed separate semantic categories. By testing toddlers with different vocabulary sizes using this paradigm, findings from this study would have shed light on the development of semantic connections as the lexicon grows larger. Unfortunately, this in-lab study also could not be completed due to COVID-19 restrictions. Instead, I turned towards the analyses of secondary data to investigate the semantic connectivity between words in toddlers' lexicon. The methodology of semantic network analyses will be discussed in the following section.

1.3.1 Semantic network analysis

A useful tool for evaluating semantic structure of the vocabulary is using network analysis, sometimes also known as graph theory. Semantic connections between multiple words can be represented using a semantic network. Each word is represented as a *node* in the network, and the connections between nodes are known as *edges*. Semantic networks for studying the structure and growth of infant vocabulary have been built using associative links (Hills, Maouene, Maouene, Sheya, & Smith, 2009; Hills, 2013; Bilson, Yoshida, Tran, Woods, & Hills, 2015), shared features (Hills et al., 2009) and speech co-occurrences (Hills, Maouene, Riordan, & Smith, 2010; Beckage, Smith, & Hills, 2011; Hills, 2013). These networks hold information about the structure of language. Speech co-occurrence networks, where words are connected by temporal co-occurrences in natural speech, represent syntagmatic links between words. The number of connections a word has in co-occurrence networks can be used as a proxy for how diverse the linguistic contexts that the word occurs in are (Hills et al., 2010). Function words that occur widely in speech (e.g., "the", "and") will have a very high number of connections, while infrequent words and words that occur in limited contexts (e.g., "ostrich") will have few connections. On the other hand, associative networks and feature networks represent paradigmatic links between words. Associative networks, typically made using norms from free association tasks, can encompass both taxonomic (e.g., “dog” and “cat”) and thematic (e.g., “dog” and “bone”) relations between word referents along with links between words’ orthography and phonology (e.g., homophones like “bee” and “be”). Meanwhile, feature networks, where nodes are linked by shared features (e.g., “has fur”, “has four legs”), highlight taxonomic relations.

We can derive network statistics for each node in the network, for example the number of connections a node has (also known as *degree*), or the centrality of a node within the network. These statistics have allowed us to study the characteristics of early-learnt words. Words that are more frequently used in speech and are semantically-related to many other words have been found to be learnt easier than less frequent and less connected words (Hills et al., 2010). Within the semantic network, some words have a very large number of connections, known as hubs. Hub words have been suggested to be responsible for short path lengths between most words in the network, as the majority of words are directly connected to hub words. If two words are not directly connected to each other but are both connected to a hub word, they will have a short path length of 2 (word1–hub–word2).

We can use semantic networks to model the structure of the lexicon, evaluating the connectivity of words that gives rise to the aforementioned priming effects. We can also use semantic networks to probe the development of the lexicon and how newly-learnt words are integrated into the lexicon. Steyvers and Tenenbaum (2005) showed that the words known by adult language users display stronger small-world network properties compared to equally-sized networks with randomly assigned edges. Small-world networks are defined by having a high clustering coefficient, high average degree and low average path lengths. This reflects strong inter-relatedness between words in the network, which has been proposed to facilitate information transfer. The small-world properties of vocabulary networks were observed across semantic networks built independently using semantic links from 3 different databases – WordNet (Miller, 1998), Roget's thesaurus (Roget, 1911) and The University of South Florida Free Association Norms (Nelson, 1999). The growth of semantic networks has been proposed to follow a pattern referred to as *preferential attachment* (Barabási & Albert, 1999). The preferential attachment model presents a rich-gets-richer scenario where the words that are more densely connected to existing words in the lexicon are learnt earlier. Subsequently, words that already have a high number of connections with other words in the lexicon will gain more connections as more nodes are added to the network. This results in a power law distribution in the network, with a small subset of nodes having a very large number of connections, while most nodes have a small number of connections. Many nodes are connected to each other via these "hub words" which have many connections. Interestingly, Hills et al. (2009) proposed that a preferential acquisition model better explained patterns in the early vocabulary growth in toddlers, over a preferential attachment model. Under the preferential acquisition model, words that are linked to many other words in the *learning environment* (as opposed to the existing lexicon) were observed to be learnt earlier.

In Chapter 6, I will explore the factors that predict earlier age-of-acquisition of early-learnt words, with a focus on lexical frequency in child-directed speech and degree of semantic connectivity in toddlers' learning environment. I will also build semantic networks from the vocabularies of toddlers aged 15 to 25 months old, investigating the pattern of connectivity between words as compared to semantic networks with random edges. If the acquisition order of early-learnt words can be linked to associative connectivity to other words, we should see higher connectivity for toddlers' vocabulary networks as compared to random networks.

1.4 Development of vocabulary in toddlers

There is extensive interest in the developmental trajectories of monolingual and bilinguals' vocabulary growth. Researchers are not only interested in the features that differ between these groups as a function of their different language experience, but also in the features that are shared across bilinguals and monolinguals. Many factors that affect the ease of word learning are expected to be shared across monolinguals and bilinguals. Words that are learnt early by toddlers tend to have higher frequency of occurrence in speech (Storkel, 2004; Hills et al., 2010), contextually diverse in speech (Hills, 2013) and shorter in length (Storkel, 2004). Among these predictors, word frequency in the input, specifically frequency of the word form and co-occurrences with its referent, is a widely-supported factor influencing word acquisition (Yu & Smith, 2007). If a word is heard more often, it is likely to be learnt earlier. This frequency effect is interesting in the context of bilingual language development. With bilingual input, word frequency is generally interlinked with the proportion of input received in each language. It is thus unsurprising that bilinguals typically have a larger vocabulary in their dominant language, which they are exposed to more often (Pearson, Fernández, Lewedeg, & Oller, 1997;

Hoff et al., 2012; Cattani et al., 2014). Studies interested in the differences between groups have compared monolingual and bilingual participants in vocabulary size (Pearson, Fernández, & Oller, 1993; De Houwer, Bornstein, & Putnick, 2014) and cognitive function (Bialystok & Martin, 2004; Van den Noort et al., 2019). Research has also suggested that bilingual language experience can influence word learning mechanisms, with bilinguals found to have lower reliance on the mutual exclusivity bias relative to monolinguals (Byers-Heinlein & Werker, 2013).

1.4.1 Bilingual language environment

The variability in bilingual environment makes bilingualism a challenging field to study. In addition to frequency of use, bilingual vocabulary growth may be modulated by the socio-cultural context. For bilinguals growing up in communities where their languages have equal social prestige and both languages are widely spoken, bilinguals may perform more similarly to monolingual peers (Smithson, Paradis, & Nicoladis, 2014). Nevertheless, even in a bilingual community such as the one found in Montréal where both English and French are widely spoken, the proportion of the two languages in infant-directed speech (when each parent is talking to their child) can differ from the proportion used in overheard speech (when parents are talking to each other) (Orena, Byers-Heinlein, & Polka, 2020). Parents may also engage in language mixing, where they switch between their two languages depending on context or preference (Byers-Heinlein, 2013). Additionally, unique challenges are faced by bilinguals where one of their languages is a less-spoken minority language. Parents can find it difficult to maintain consistent input in the minority language for their child. Even when English-Irish bilingual parents report that they always speak Irish (the minority language) to their child, many families report mixed use of Irish and English (the majority language) within the household (O'Toole & Hickey, 2017). Bilingual families in the United Kingdom face a similar scenario. While there are many bilingual households in the UK, the home language(s) (henceforth Additional Language, or AL) varies greatly. According to the England and Wales 2021 Census, 7.1% of the population spoke English fluently but did not consider it their main language (Office for National Statistics, 2022). The two 5 main languages of residents in the 2021 Census, listed in decreasing number of residents, were Polish (1.1%), Romanian (0.8%), Panjabi, Urdu and Portuguese. It is also not uncommon to have bilingual parents who do not share a main language or whose only shared language is English. For 1.9% of England and Wales households in the 2021 Census, the main language spoken at home differed between generations, and an additional 2.2% of households had the main language spoken differ between partnerships (Office for National Statistics, 2022). The diverse scenarios as a result of varying language status, positive or negative view of bilingualism, parents' native language and parents' shared language result in different language profiles among bilinguals.

Therefore, it is important when conducting research with bilingual populations to be aware of possible variability in the sample. Questionnaires that probe aspects of bilingual language exposure in infants and toddlers have been put forward by Bosch and Sebastián-Gallés (2001) and Byers-Heinlein (2013), among others. These questionnaires provide researchers with a method to quantify task-relevant aspects of bilingual language exposure. There have also been explorations of more complex models for defining patterns of bilingualism that combine features of both the categorical approach (e.g. monolingual vs bilingual) and continuous approach (e.g. L2 exposure of 0% to 100%, where monolinguals lie on the extreme ends of the spectrum). Kremin and Byers-Heinlein (2020) suggested the use of factor mixture models (where category members can vary in score within the category) or grade-of-membership models (where ambiguous cases can occur across multiple categories). With these available options, the most appropriate method for defining bilinguals would ultimately depend on the research question. Kremin and Byers-Heinlein highlighted the importance of transparency in reporting the characteristics of each study's bilingual samples, for the sake of better understanding and reproducibility in the field of bilingual research.

In Chapter 2, I will compare the vocabulary sizes of monolingual and bilingual toddlers, looking at the effect of language exposure on English vocabulary size, conceptual vocabulary size and total vocabulary size. Demographic and language environment information of the bilingual sample will be reported in Appendix A.

1.5 Overview of upcoming chapters

In this thesis, I will show evidence of how bilinguals' two languages interact with each other, focusing on phonological and semantic links between words. I discuss how cross-linguistic interactions can result in vocabulary learning trajectories that differ from monolinguals in vocabulary size and composition. The majority of chapters will present analyses conducted on bilingual and monolingual data collected using vocabulary questionnaires. These toddlers grew up in the UK, where the use of English is dominant in the community. Bilingual toddlers in this sample therefore represent a subset of bilinguals who hear their additional language mostly in the home, with limited exposure to the additional language in the community. I will start by comparing the vocabulary sizes of these bilingual and monolingual toddlers in Chapter 2.

I will then move on to explore effects of phonological similarity in the lexicon. Before evaluating the effect of cognates on word learning and recognition, I will review the methodology of defining cognate status using phonological similarity. I evaluate two established algorithms for phonological similarity in Chapter 3, highlighting their strengths and weaknesses in comparison to a behavioural measure of cognateness. The algorithm with better performance will be used for cognate classification in subsequent analyses on cognates in this thesis. Then, in Chapter 4, I evaluate the use of a computational model of spoken word recognition for predicting monolingual participants' recognition of unfamiliar foreign words, using the behavioural data described Chapter 3. I then return to the vocabulary questionnaire data to study the effect of cognates on vocabulary trajectories, using cross-sectional data for the preliminary analyses and supporting it with complementary findings from longitudinal data. This will be presented in Chapter 5. As previously-mentioned, two other chapters covering in-lab experimental studies were initially planned to be included in this thesis, but difficulty in recruiting the necessary sample during the COVID-19 pandemic resulted in a shift towards studies conducted online. The first in-lab study would have investigated bilingual toddlers' activation of their two languages during spoken word recognition, looking at how phonological competitors in both the test language and non-test language may impact target recognition using a phonological priming paradigm. It would have supplemented the research in this thesis by providing insight into the effect of phonological connections not just on word learning but also on word recognition.

The final part of this thesis will focus on semantic connections between words. The second in-lab study would have probed the development of semantic connections as toddlers grew to know more words. It would have supplemented the research topic of semantic connectivity in the lexicon, providing insight into the activation and inhibition of semantic connections in response to task demands. However, as with the previously-mentioned in-lab study, this study also could not be completed due to lockdown restrictions. Instead, I conducted semantic network analyses on secondary data of monolingual and bilingual toddlers' vocabulary knowledge. In Chapter 6, I evaluate various predictors for toddlers' word acquisition order, focusing on lexical frequency and semantic connectivity. I will then explore the semantic connections between words in toddlers' vocabularies, investigating whether toddlers' word learning trajectories follow patterns that reflect sensitivity to semantic connections in the learning environment.

By investigating both phonological and semantic connections between words in bilingual toddlers' lexicons, I aim to broaden our knowledge of the factors guiding bilingual toddlers' vocabulary learning.

Chapter 2

Comparing bilingual toddlers' vocabulary size with monolingual toddlers' vocabulary size

2.1 Background

Researchers, medical practitioners and parents alike have long been interested in the question of whether bilinguals' vocabulary growth is comparable to that of monolinguals. For bilingual toddlers living in communities that predominantly speak one language (e.g., the UK, which uses English), it is particularly important

---

This chapter includes elements that have been published with the following reference: Siow, S., Lepadatu, I., Gillen, N., & Plunkett, K. (2022). UK bilingual toddlers show a lag in vocabulary size relative to monolinguals in both comprehension and production. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 44, No. 44). SS was responsible for data analysis, interpreting results and writing the manuscript.

for them to acquire the community language which would be widely used for communication outside the home and schooling. The literature on bilinguals' vocabulary size has generally found a bilingual delay when comparing vocabulary in a single language between monolinguals and bilinguals. When tested using a receptive vocabulary task (Peabody Picture Vocabulary Test; PPVT-III; L. Dunn & Dunn, 1997), significant differences have been found between monolingual and bilingual groups in receptive vocabulary in children (3–10 years old) (Bialystok, Luk, Peets, & Sujin, 2010) and adults (Bialystok & Luk, 2012), with bilinguals having smaller vocabulary sizes than monolinguals. With school-age children, Yan and Nicoladis (2009) found that while comprehension (as tested using the PPVT-III) was comparable, school-age bilinguals performed significantly poorer in a production task (picture naming) compared to monolingual peers.

When considering younger bilinguals, a bilingual vocabulary delay in production (tested using the Expressive One-Word Picture Vocabulary Test; Brownell, 2000) has also been found in young children aged between 2.5 and 5 years old when compared against monolingual peers (Hoff & Ribot, 2017). This group difference was also found by Vagh, Pan, and Mancilla-Martinez (2009) in 24 to 36 month olds when using the MacArthur-Bates CDI: Words and Sentences to measure productive vocabulary size. Cattani et al. (2014) tested toddlers (22 to 30 months old) using a receptive vocabulary task (The British Picture Vocabulary Scale III; L. M. Dunn & Dunn, 2009), an expressive vocabulary task (picture naming), and parent-reported vocabulary in comprehension and production (Oxford CDI; Hamilton, Plunkett, & Schafer, 2000). They found that Spanish-English bilingual toddlers performed worse than English monolingual toddlers when tested using these tasks in English. Interestingly, De Houwer et al. (2014) found no significant difference in the receptive vocabulary sizes in Dutch (tested using the Dutch adaptation of the CDI; Zink & Lejaegere, 2002) of monolingual toddlers learning Dutch and bilingual toddlers learning Dutch and French at 13 months, but monolinguals knew significantly more Dutch words at 20 months, suggesting an age-related change in vocabulary growth. Smaller English vocabulary sizes for 24-month-old bilinguals relative to monolinguals in both comprehension and production have also been found by Floccia et al. (2018) for UK bilingual toddlers.

2.1.1 Academic achievement

For bilingual toddlers living in communities that predominantly speak one language (e.g., the UK, in which the main language is English), it is particularly important for them to acquire the community language which would be widely used for communication outside the home and schooling. Language proficiency in the majority language can have repercussions for school achievement. A study by Howard et al. (2014) showed that Spanish-speaking bilingual children's English vocabulary size in spoken production (as tested using a picture naming task) is positively associated with their English reading proficiency, even after accounting for the effect of socioeconomic status and amount of English exposure. A report by Strand, Malmberg, and Hall (2015) analysing the England National Pupil Database in 2013 indicated that the percentage of students in England classified to be learning English as an additional language (EAL) was 16.2%. As a group, EAL students were identified by Strand et al. to have lower rates of academic achievement compared to students with English as their first language when tested at the end of their first year of schooling. However, this lag decreased over the years of schooling, with EAL students catching up to their peers by age 16. Research has suggested that within-group differences in school-age language outcomes can be predicted by language development in infancy. In monolinguals, larger vocabulary size and faster speed of word recognition tested at 25 months of age have been linked to better expressive vocabulary, IQ and working memory at 8 years old (Marchman & Fernald, 2008). Studying the early vocabulary development of children in their first three years of life, when their early language skills are rapidly developing, can help us better understand the potential sources of divergences for EAL students.

2.1.2 Effect of language exposure

The size of the vocabulary difference between monolinguals and bilinguals is also dependent on the amount of exposure bilinguals receive for the tested language. Vocabulary size in a single language has been found to be positively correlated with the relative amount of exposure the child has to that language (Pearson et al., 1997; Hoff et al., 2012; Cattani et al., 2014). English-dominant bilingual toddlers (i.e., toddlers who hear more English than their other language in their day-to-day lives) have been found to display larger English vocabulary than Spanish-dominant bilinguals, using evidence from 8–30 month olds (Pearson et al., 1993) and 24–36 month olds (Vagh et al., 2009). Further supporting the effect of language exposure, Pearson et al. (1993) also found that while Spanish-dominant bilinguals had smaller English vocabulary, they had larger Spanish vocabulary than English-dominant peers. It is particularly important to be aware of the language exposure effects when comparing bilinguals to monolinguals in a single language, as it can significantly affect the size of any observed vocabulary gap. Notably, Cattani et al. (2014) found that 2.5 year old bilingual toddlers with at least 60% English exposure performed equally well as English monolingual peers on various language measures in English.

2.1.3 Single language vs total vs conceptual vocabulary

When studying vocabulary growth in bilinguals, the method of calculating vocabulary size is important as it can produce varied results. Researchers have used several measures for vocabulary, the most common being single language size, total vocabulary size and conceptual vocabulary size. Single vocabulary sizes focuses on vocabulary known in one language, for example the community language or the minority language. Total vocabulary size sums the vocabulary sizes in both languages. Conceptual vocabulary is defined by summing the number of concepts known by the child. A child is said to know a concept if they understand the word in one language or both. Conceptual scoring has been noted to bring school-age bilingual's vocabulary into normal monolingual range for both comprehension and production (Gross, Buac, & Kaushanskaya, 2014). Bilingual toddlers have been found to have smaller vocabularies than monolinguals when comparing single language vocabulary, but comparable or even larger vocabularies when comparing total vocabulary and conceptual vocabulary (Pearson et al., 1993, 1997). The appropriate method of vocabulary size calculation would therefore depend on the intention of the comparison. The evaluation of single language vocabulary size (e.g., of the majority language) may be useful when investigating later language and academic outcomes. For clinical judgements of language delay, a conceptual or total vocabulary would provide a more reliable estimate. In this chapter, I compare bilingual and monolingual toddlers using all three methods of calculating vocabulary size.

In addition to these calculations of vocabulary sizes, we can also explore the interconnectivity of bilinguals' two languages by studying the overlap in bilinguals' vocabulary in their two languages. Words that share the same meaning across languages are known as *translation equivalents*. When a child knows both words in a translation equivalent pair, they are considered to know a *doublet*. This is in contrast to translation equivalent pairs where the child knows only one word of the pair, otherwise referred to as a *singlet*. Research with bilingual children's early vocabulary has found doublets in the receptive vocabulary of toddlers as young as 13 months old (De Houwer, Bornstein, & De Coster, 2006) and in productive vocabulary by 16 months old (Legacy et al., 2017).

2.1.4 The Present Study

This study investigates whether bilingual toddlers growing up in the UK have comparable or smaller vocabulary sizes compared to monolinguals of the same age. As the UK is a predominantly English-speaking community, the development of English proficiency is important for both monolingual and bilingual toddlers' long-term communicative and academic outcomes. I am also interested in the extent to which the degree of English exposure a child receives influences their English vocabulary size. Additionally, I investigate whether bilinguals and monolinguals have comparable vocabulary sizes when measured using conceptual vocabulary and total vocabulary. Finally, I also calculate the proportion of doublets, to study the interactions between the two languages.

I compared vocabulary acquisition trajectories between British monolinguals and bilinguals growing up in the UK aged 12 to 36 months old, comparing cross-sectional data collected using vocabulary questionnaires. I obtained parent-reported data on both word comprehension and production for each child, to study toddlers' parallel growth in comprehension and production. I predicted that bilinguals will have smaller vocabulary sizes in English than monolinguals, with the difference largest for AL-dominant bilinguals and smallest for English-dominant bilinguals. On the other hand, I predicted that all groups will have similar conceptual vocabulary sizes, with no significant differences between bilinguals and monolinguals after controlling for age and mother's education level.

2.2 Methods

2.2.1 Participants

Bilingual

The full sample consisted of 12 to 36-month-old bilingual toddlers (N = 622; N female = 321, N male = 300, N other = 1) (age 12.0–36.0, mean 23.7 months) growing up in the UK with English and one additional language (AL) (Dutch, French, German, Italian, Polish, Portuguese or Spanish), with data collected between 2020 and 2022. Participants were recruited via advertisements on social media (restricted to families living in England) and via email to families who had signed up to the lab's database. Informed consent was obtained from parents filling in the questionnaire. Bilingual families were offered the option of a £5 Amazon voucher or a child-sized t-shirt as remuneration.

I required participants to hear at least 20% of each language. I excluded participants who reported hearing 10% or more of a third language. I required at least one parent to be a native speaker of the minority language. I excluded participants who were reported to be premature by 6 weeks or more, and those who reported hearing problems or diagnosed language delay. To obtain reliable estimates of vocabulary size, I excluded an additional 42 parents who expressed uncertainty about their ability to report their child's English vocabulary (e.g., due to not speaking English at home). Further details about the language background of bilingual toddlers in the sample can also be found in Appendix A.

For each family in the sample, at least one parent was a native speaker of the AL – 220 reported that both parents were native speaker of the AL (this included those who reported to be natively bilingual); 364 families reported that one parent was a native speaker of the AL and one parent was a native speaker of English; 38 families had one AL-native parent and one parent who was native in a third language. The parent native in the AL was more commonly the mother – 535 mothers were reported to be native AL speakers, 32 native bilingual speakers, 48 native English speakers and 7 native in another language. In contrast, 255 fathers were reported to be native AL speakers, 17 bilingual speakers, 321 English native speakers and 29 native in another language. I also required at least one parent to have fluent English proficiency (self-rated proficiency of 7 or higher out of 10).

On average, UK bilingual toddlers were slightly skewed to have more exposure to English than to their AL in their daily lives, with mean English exposure for the overall sample at 54.3%. I split bilinguals into three groups based on their reported English exposure, with thresholds selected to divide the sample into similarly-sized groups: 227 English-dominant bilinguals (65-80% English exposure, mean = 73.2, SD = 4.99), 211 Balanced bilinguals (41-64% English exposure, mean = 54.8, SD = 5.27) and 185 AL-dominant bilinguals (20-40% English exposure, mean = 30.7, SD = 7.39).

Monolingual

The monolingual sample consisted of British English monolinguals aged between 12 and 36 months (N = 239; N female = 89, N male = 150) (age 12.3–36.0, mean 24.0

months), with data also collected between 2020 and 2022.

Monolingual participants completed the questionnaires reported in this paper as part of other studies. A subset (N = 180) participated in an online study testing the utility of a touchscreen receptive vocabulary task Gillen et al. (2021). These families participated without remuneration. All questionnaires reported in the present study were completed prior to any additional tasks. The remaining monolingual participants (N = 77) completed the questionnaires as part of their participation in a lab-based experiment. These participants were offered £5 for travel expenses and a child-sized t-shirt as appreciation for their participation in the in-person study. Questionnaires were sent to parents via email to be completed online in the comfort for their own home. As such, we do not expect that participation in either of the two aforementioned studies would have any significant impact on the analyses conducted in this paper, as the order of the online questionnaires was identical to that of bilingual participants who completed the full study online.

2.2.2 Procedure

Questionnaires were administered online using Qualtrics (Qualtrics, 2022). Parents received a link to the questionnaires that they could complete at home. The questionnaires could be paused and resumed using the same link, but would time out if not completed within one week.

Demographics questionnaire

All parents first answered some questions relating to their demographics (toddlers' date of birth, each parent's highest level of education, any languages spoken at home besides English, and whether the toddler had any diagnosed language delay, hearing problems or was born premature).

I used mother's highest education level as a proxy for socioeconomic status. Only entries where information on mother's education level was available were included in the analysis. Education level was converted into a numerical score, with 0 – no qualifications; 1 – Left school at 16 with GCSE or equivalent; 2 – Left school at 18 with A-Levels or equivalent; 3 – University degree or equivalent. Overall, mothers' educational level in the sample was high, with 90.5% of mothers in the bilingual sample and 93.7% of mothers in the monolingual sample having a University degree or equivalent.

Language exposure questionnaire

Bilingual parents (who reported that their child had regular exposure to English and one additional language) went on to complete a simplified version of the Language Exposure Questionnaire (LEQ) developed by (Bosch & Sebastián-Gallés, 2001). The LEQ is a parental report questionnaire used to obtain a summary of each toddlers' language environment. It includes questions about each parent's native language, their language proficiency, which languages(s) they usually use when speaking to the toddler, whether or not the toddler is attending nursery (and the language used as nursery), and whether the toddler has spent time immersed in a country where their AL is widely spoken. The list of questions included in the LEQ can be found in Appendix A. To obtain a quantitative metric for toddlers' language exposure, we also asked parents to give an estimate of the percentage of English, AL and any third language that their child is exposed to in their daily life overall as part of the LEQ. Parents were also asked about the percentages of these languages heard specifically in the home. Monolingual parents (who reported that their child had no consistent exposure to any other languages besides English) were not administered the LEQ.

Following the demographics questions and LEQ (if applicable), parents were directed to the appropriate version of the vocabulary questionnaires according to their toddler's language background.

Vocabulary questionnaire

Data on vocabulary knowledge in English, for both the bilingual and monolingual groups, was collected using the Oxford Communicative Development Inventory (CDI) (Hamilton et al., 2000), which is a questionnaire containing a list of words commonly known to British toddlers. Parents indicated for each word whether their child understands and says, understands but does not say, or does not understand the word. The utility of CDIs to evaluate vocabulary development in toddlers has been supported by studies showing good congruence between parent-reported vocabulary and toddlers' performance on vocabulary tasks for both monolinguals (Styles & Plunkett, 2009; Gillen et al., 2021) and bilinguals (Marchman &

Martínez-Sussmann, 2002; Vagh et al., 2009). Vocabulary data for the monolingual sample was collected using the Oxford CDI (418 words). The full list of Oxford CDI words can be found in Appendix B. Parents of bilingual toddlers completed the Oxford CDI and also an adaptation of the Oxford CDI in their AL (also 418 words). These adaptations were created by working with native speakers of each AL, who translated the Oxford CDI and replaced words that were not relevant to the target language – for example, "penny" was replaced with its closest equivalent "coin" in most languages. If two Oxford CDI entries had the same word as their translations (e.g., English "clock" and "watch" are both translated as "reloj" in Spanish), a replacement word of the same category and similar age-of-acquisition was selected to be included in the adaptation, so that words were not duplicated in the CDI. If an English word had multiple translations, for example due to dialectal differences, I listed the different translations as one entry separated by forward slashes. I also compared the translations to normed adaptations of the MacArthur-Bates CDI in those languages, using the same words if possible – to given an example, "lorry/truck" was listed as "Lastwagen / Laster" in our German CDI, following the equivalent item in FRAKIS (Szagun, Stumper, & Schramm, 2009). The list of overlapping concepts across the Oxford CDI and all adaptations used in this thesis can be found in Appendix C, with the entries listed exactly as they would have been shown in the questionnaires.

For all analyses reported in this chapter, I used only the concepts that overlap across all the adaptations (365 out of 418 words). While normed versions of the CDI exist in these languages, they vary considerably in length and also have variable amounts of overlapping concepts with the Oxford CDI. I chose to use adaptations of the Oxford CDI as this allowed me to have a high level of conceptual overlap for analyses of conceptual vocabulary size. English vocabulary sizes, conceptual vocabulary sizes and total vocabulary sizes were calculated using the 365 concepts that overlap across all CDIs used in this study. A monolingual child was coded as knowing a concept if they knew the English word for the concept. A bilingual child was coded as knowing a concept if they knew the English word, the word in their other language, or both.

2.2.3 Sample subset for vocabulary in comprehension

The Oxford CDI, like other vocabulary questionnaires that use a standardised list to measure vocabulary development across different ages, is prone towards floor and ceiling effects. The CDI is a finite list of words formulated to compare vocabulary growth over a narrow age range, and does not claim to be a comprehensive list of all words that toddlers can possibly know. Toddlers, especially those who have larger vocabularies, are likely to also know words that are not included in the CDI. The finite list means that we cannot differentiate between a child who knows all 418 words in the CDI and no other words, and a child who knows all 418 words and an additional 400. Ceiling effects are problematic in analyses of vocabulary growth, as it generates a plateau of scores at the older ages that is not reflective of the actual developmental trajectory. It is therefore important to specify a cut-off that minimises the possibility of scores reaching ceiling. Floor effects create a similar problem, creating a plateau at the youngest ages. Very young infants are unlikely to know many words, and in fact may not know any words. Importantly, as toddlers typically learn to produce words later than they learn to understand them, the age at which toddlers' vocabulary sizes hit the floor and the ceiling are likely to differ for comprehension and production, thus requiring different cut-offs.

To avoid floor and ceiling effects on vocabulary size comparisons, analyses on vocabulary size in comprehension used a subset of participants from the original 12 to 36-month-old sample. I defined the ceiling as 90% of the maximum conceptual vocabulary size as calculated from overlapping concepts in our questionnaires, and the floor as 10% of the maximum conceptual vocabulary size. I excluded age groups that had median vocabulary sizes in comprehension smaller than the floor value and those that had median vocabulary sizes larger than the ceiling value. The bilingual group and the monolingual group had their thresholds at different ages – for the bilingual group, the lower cut-off was 12 months and the upper cut-off was 26 months; for the monolingual group, the lower threshold was 14 months and the upper threshold was 25 months. I applied the more conservative age range for both groups, leaving a sample subset of 14 to 25-month-old toddlers for analyses of vocabulary size in comprehension. This subset included 141 monolingual toddlers and 297 bilingual toddlers (106 English-dominant, 95 Balanced and 96 AL-dominant).

2.2.4 Sample subset for vocabulary in production

I similarly aimed to avoid floor and ceiling effects for vocabulary size in production. As with comprehension, the ceiling was defined as 90% of the maximum conceptual vocabulary size, and the floor as 10% of the maximum conceptual vocabulary size. For the bilingual group, the lower cut-off was 20 months and the upper cut-off was 36 months; for the monolingual group, the lower threshold was 19 months and the upper threshold was 29 months. Again, I applied the more conservative age range for both groups. The sample subset for analyses of vocabulary size in production was 20 to 29-months-old. This subset included 151 monolingual toddlers and 267 bilingual toddlers (99 English-dominant, 92 Balanced and 76 AL-dominant).

2.3 Results

I ran linear regressions (separately for comprehension and production) with vocabulary size as the dependent variable, language exposure group as the predictor, and age, gender and mother's highest education level (numerical score) as covariates. Age was centered on the mean and scaled by standard deviation for each vocabulary size measure separately. As there was only one participant with gender classified as "Other", they were excluded from further analyses. The three bilingual groups (English-dominant, Balanced and AL-dominant) were contrasted against the reference level of monolinguals. This analysis was done for both English vocabulary size and conceptual vocabulary size separately. The model is defined in R (R Core Team, 2013) as below:

```
lm(vocabulary_size ~ age + gender +
   mother_highest_education + group)
```

2.3.1 Comprehension (14 to 25-month-old)

The relationship between age, language group and English vocabulary size in comprehension is visualised in Figure 2.1. We see the expected strong positive trend of vocabulary size growth with age. As predicted, we also observe a difference between the vocabulary trajectories of monolinguals and bilinguals, with bilinguals of all three levels of language exposure having smaller vocabulary sizes in English compared to monolinguals of the same age.

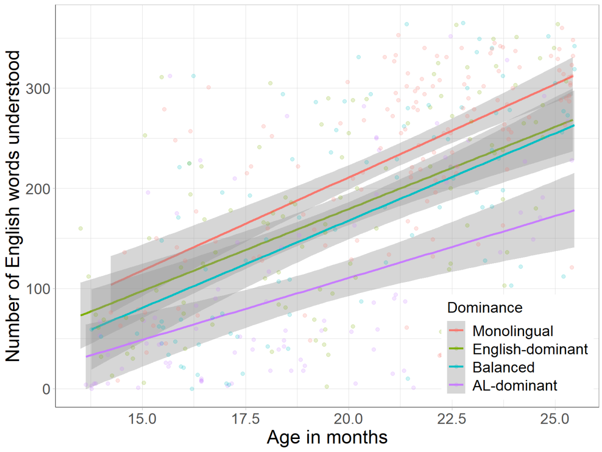

Figure 2.1: Scatterplot of English vocabulary size in comprehension against toddler's age, split by language dominance groups.

On average, the bilingual sample had smaller vocabulary sizes than same-age monolinguals for English vocabulary in comprehension ($t = -6.770$, $p < .001$). When bilinguals are split by dominance groups, we see the expected trend for English vocabulary size across groups. All three groups had significantly smaller vocabulary size compared to monolinguals (Table 2.1). As shown by the model estimates, on average English-dominant bilinguals knew 35 fewer English words than monolinguals (out of 365 concepts), Balanced bilinguals knew 46 fewer words and AL-dominant bilinguals knew 99 fewer words. Tukey HSD post-hoc tests confirmed that the 3 bilingual groups showed the expected pattern in English vocabulary size, with AL-dominant bilinguals having smaller English vocabulary sizes in comprehension than both balanced (t = -4.549, $p < .001$) and English-dominant bilinguals (t = -5.567, $p < .001$). While balanced bilinguals showed a trend of having smaller English vocabulary sizes than English-dominant bilinguals, the difference was not significant (t = -0.907, p = .801).

Table 2.1: Linear model for English vocabulary size in comprehension, with age and language dominance as predictors (Monolingual is reference level).

Predictor	Estimate	Std Error	t	p
(Intercept)	285.0	26.8	10.6	$<$.001
Age	106.0	7.67	13.8	$<$.001
Gender	-14.1	7.79	-1.81	.0707
Mother's education	1.53	8.82	0.173	.862
English-dominant	-36.8	10.6	-3.46	$<$.001
Balanced	-45.2	10.8	-4.36	$<$.001
AL-dominant	-100.7	11.0	-9.12	$<$.001

I then studied the relationship between age, language group and conceptual vocabulary size in comprehension. In Figure 2.2, which visualises the relationship for conceptual vocabulary in comprehension, we see that the difference between bilinguals with different levels of language exposure (as seen in the previous figure of English vocabulary comprehension) has largely disappeared. The difference between the monolingual group and the bilingual groups has also reduced.

The bilingual sample had comparable vocabulary sizes in comprehension with same-age monolinguals when vocabulary size was calculated using conceptual vocabulary (t = -1.39, p = .166). All three bilingual groups did not differ significantly

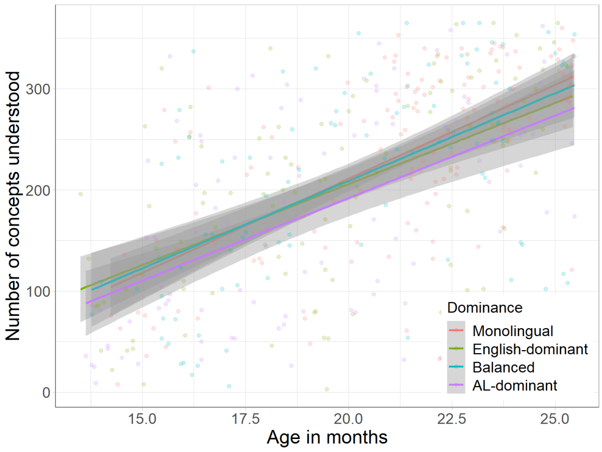

Figure 2.2: Scatterplot of conceptual vocabulary size in comprehension against toddler's age, split by language dominance groups.

Table 2.2: Linear model for conceptual vocabulary size in comprehension, with age and language dominance as predictors (Monolingual is reference level).

Predictor	Estimate	Std Error	t	p
(Intercept)	278.7	25.9	10.7	<.001
Age	111.3	7.43	15.0	<.001
Gender	-14.4	7.55	-1.91	.0572
Mother's education	4.60	8.55	0.538	.591
English-dominant	-9.18	10.3	-0.891	.373
Balanced	-5.04	10.5	-0.481	.631
AL-dominant	-21.3	10.7	-1.99	.0474

in conceptual vocabulary size compared to monolinguals (Table 2.2). Tukey HSD post-hoc tests also showed no significant differences between the three bilingual groups.

Analyses of the proportion of doublets in bilingual toddlers' vocabulary showed that bilingual toddlers have a large number of words that they understand in both languages. While the exact ratio varies greatly between participants (range 0–

99.7%), on average participants understood both translation equivalents of 47.3% (SD = 24.8%) of the concepts they knew. As a result of this high rate of doublets, bilingual toddlers know a large number of words when their two languages are added together, exceeding the vocabulary size of monolinguals of the same age (Figure 2.3, Table 2.3). Interestingly, Tukey HSD post-hoc tests revealed a significant difference between balanced and AL-dominant bilinguals, with balanced bilinguals having significantly larger total vocabulary size (t = 3.192, p = .00806). Balanced bilinguals also had a higher proportion of doublets in their vocabulary than AL-dominant bilinguals (t = 5.327, $p < .001$). Differences between balanced bilinguals and English-dominant bilinguals were not significant for total vocabulary size and proportion of doublets.

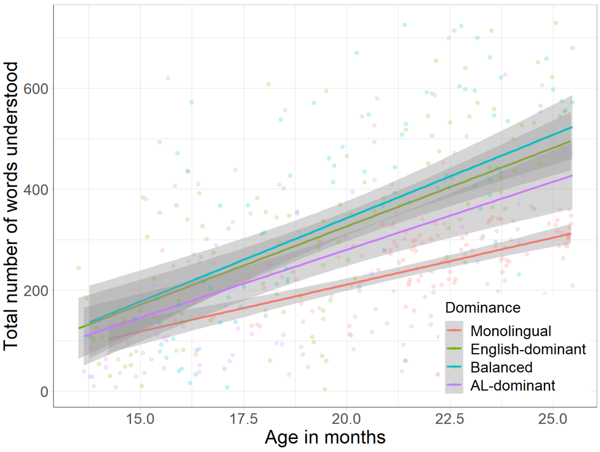

Figure 2.3: Scatterplot of total vocabulary size in comprehension against toddler's age, split by language dominance groups.

Table 2.3: Linear model for total vocabulary size in comprehension, with age and language dominance as predictors (Monolingual is reference level).

Predictor	Estimate	Std Error	t	p
(Intercept)	287.6	44.66	6.440	$<$.001
Age	176.3	12.78	13.79	$<$.001
Gender	-22.72	12.99	-1.749	.0811
Mother's education	13.25	14.71	0.901	.368
English-dominant	120.1	17.73	6.775	$<$.001
Balanced	141.1	18.03	7.826	$<$.001
AL-dominant	78.47	18.42	4.261	$<$.001

2.3.2 Production (20 to 29-month-old)

Differences between monolinguals and bilinguals were also found in vocabulary size for production (Figure 2.4). Bilinguals produced less English words than monolinguals (t = -6.979, $p < .001$), with the difference from monolinguals largest for AL-dominant bilinguals and smallest for English-dominant bilinguals (Table 2.4). Again, Tukey HSD tests showed significant differences between AL-dominant bilinguals and English-dominant bilinguals (t = -4.831, $p < .001$), but the difference between balanced bilinguals and English-dominant bilinguals was not significant (t = -1.369, p = .518).

As with comprehension, conceptual vocabulary scoring reduced the difference between monolinguals and bilinguals. However, in contrast to the results with comprehension, there remained a difference between groups for conceptual vocabulary production, as seen in Figure 2.5. This difference was significant between monolinguals and AL-dominant bilinguals, with AL-dominant bilinguals producing fewer concepts than monolinguals. The difference was only marginally signifi-

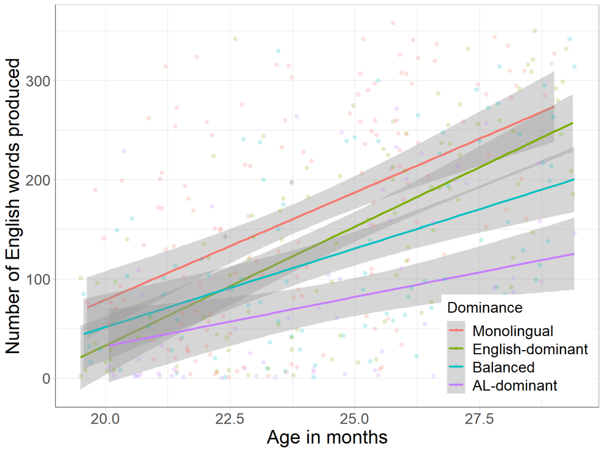

Figure 2.4: Scatterplot of English vocabulary size in production against toddler's age, split by language dominance groups.

Table 2.4: Linear model for English vocabulary size in production, with age and language dominance as predictors (Monolingual is reference level).

Predictor	Estimate	Std Error	t	p
(Intercept)	194.7	29.31	6.64	$<$.001
Age	121.0	10.11	11.97	$<$.001
Gender	-31.8	8.42	-3.77	$<$.001
Mother's education	-2.952	9.89	-0.298	.766
English-dominant	-38.5	11.24	-3.42	.0042
Balanced	-55.4	11.28	-4.91	$<$.001
AL-dominant	-101.5	11.97	-8.48	$<$.001

cant between monolinguals and English-dominant bilinguals, and between monolinguals and balanced bilinguals (Table 2.5). Differences between the three bilingual groups were not significant when tested using Tukey HSD tests.

As with comprehension, bilingual toddlers have a large number of doublets in their vocabulary, albeit lower than the values seen in comprehension. Again the proportion varies (range 0–89.8), but on average toddlers in the sample produced

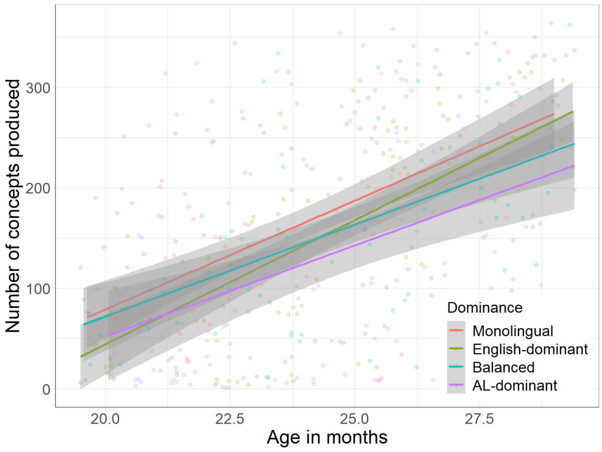

Figure 2.5: Scatterplot of conceptual vocabulary size in production against toddler's age, split by language dominance groups.

Table 2.5: Linear model for conceptual vocabulary size in production, with age and language dominance as predictors (Monolingual is reference level).

Predictor	Estimate	Std Error	t	p
(Intercept)	188.6	30.32	6.222	<.001
Age	137.3	10.46	13.12	<.001
Gender	-34.01	8.71	-3.904	<.001
Mother's education	-0.196	10.23	-0.019	.985
English-dominant	-26.06	11.63	-2.241	.0256
Balanced	-25.59	11.67	-2.194	.0288
AL-dominant	-44.65	12.38	-3.606	<.001

the translation equivalents in both languages of 31.6% (SD = 22.0%) of the concepts they produced.

In contrast to total vocabulary size in comprehension, where bilinguals had larger vocabulary sizes than monolinguals, total vocabulary size in production had a more complicated pattern. Only balanced bilinguals had significantly larger total vocabulary size in production than monolinguals. AL-dominant bilinguals had similar total vocabulary size in production with monolinguals (Figure 2.6, Table 2.6). Differences in total vocabulary between the three bilingual groups were not significant when tested using Tukey HSD tests. However, balanced bilinguals produced a significantly higher proportion of doublets than both English-dominant bilinguals (t = 3.566, $p < .001$) and AL-dominant bilinguals (t = 3.538, $p < .001$).

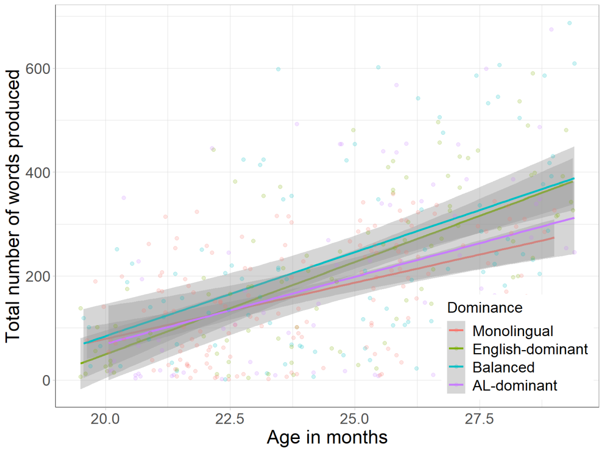

Figure 2.6: Scatterplot of total vocabulary size in production against toddler's age, split by language dominance groups.

Table 2.6: Linear model for total vocabulary size in production, with age and language dominance as predictors (Monolingual is reference level).

Predictor	Estimate	Std Error	t	p
(Intercept)	188.3	43.74	4.306	<.001
Age	188.4	15.09	12.49	<.001
Gender	-40.9	12.57	-3.254	.00123
Mother's education	1.935	14.76	0.131	.896
English-dominant	21.70	16.78	1.294	.197
Balanced	46.02	16.83	2.735	.00652
AL-dominant	2.288	17.86	0.128	.898

2.4 Discussion

2.4.1 Effects of language input

In this chapter, I present findings regarding the vocabulary growth of a sample of UK bilingual and monolingual toddlers that are convergent with previous findings in the literature (Vagh et al., 2009; Cattani et al., 2014; Hoff & Ribot, 2017). For English vocabulary size, there was a trend for bilinguals' vocabulary size to be smaller than same-age monolinguals, with the difference significant in both comprehension and production. This difference was modulated by the amount of English exposure received by the child, with English-dominant bilinguals being most similar to monolinguals (but still significantly smaller in English vocabulary size) and AL-dominant bilinguals having the smallest English vocabulary size. The relationship between the percentage of English exposure and vocabulary size in English highlights the importance of language input in vocabulary learning. Words that are learnt earlier have been associated with higher frequency in the language input (Goodman, Dale, & Li, 2008; Hills et al., 2010; Hills, 2013; Braginsky, Yurovsky, Marchman, & Frank, 2019). It is therefore unsurprising that toddlers who hear less of a language also have slower vocabulary growth in that language.

Interestingly, the differences in English vocabulary were not significant between English-dominant and balanced bilinguals in both comprehension and production – in fact these two groups unexpectedly showed very similar patterns in vocabulary development despite their differences in language exposure. The group of balanced bilinguals have a mean English exposure of 54.8%, while the

English-dominant bilinguals had a mean English exposure of 73.2%. The disportionally small difference in English vocabulary size may point towards the possibility of that the effect of language exposure on vocabulary trajectories may not be linear as I previously assumed. It is possible that vocabulary gains from increased exposure to a particular language may be smaller when language input already is skewed towards that language. Further investigations of the effect of language input on bilingual language development can help us better understand the amount of language input required to support the acquisition of each language.

2.4.2 High conceptual overlap in bilingual vocabulary

It is also important to remember that for bilinguals, their vocabulary size is split between two languages. Consistent with this, I found that the bilingual toddlers in the sample understood similar numbers of concepts compared to monolinguals of the same age. Strikingly, when comparing total vocabulary size in comprehension, bilingual toddlers showed larger vocabulary sizes than monolinguals. The larger total vocabulary sizes of bilinguals toddlers relative to monolinguals can be attributed to a high number of words that bilingual toddlers understand in both languages. As a result of receiving language input in two languages, bilingual toddlers acquire words in both languages which overlap in meaning. This parallel development of their two languages is important for the communicative needs of bilingual toddlers, who may need different languages to communicate effectively with the people in their lives. This hypothesis is supported by the observation that balanced bilinguals had larger total vocabulary and a higher proportion of doublets in their vocabulary than AL-dominant bilinguals.

The acquisition of doublets may be additionally supported by cross-linguistic phonological overlap. Translation equivalents that also share phonological properties (e.g. English “fish” and Dutch “vis”) are known as cognates. The high number of similar sounds that cognates share may help the acquisition of new words, particularly in production. The activation of a cognate word in one language has been proposed to simultaneously activate its translation equivalent in the other language, which in turn feeds activation back to the target word via its shared phonology (Costa et al., 2000). This theory was used by Costa et al. to explain faster picture naming for cognates than non-cognates. The effect of cognates on bilingual toddlers’ comprehension and production trajectories will be explored in Chapter 5 of this thesis. In addition to cross-linguistic phonological overlap, there may also be a facilitatory effect of phonological neighbours. Sosa and Stoel-Gammon (2012) found that words with higher phonological neighbourhood density were produced by toddlers more consistently and closer to adult pronunciations than words with lower phonological neighbourhood density. They suggested that a high number of neighbours may prompt stronger and more detailed phonological representations to be formed during learning, or that the retrieval of these words may be facilitated by high activation from neighbours feeding into the target word. Theories of non-selective bilingual lexical access suggest that the size of phonological neighbourhoods are larger for bilinguals due to the simultaneous activation of words in both languages. Particularly, cognates would activate a large neighbourhood of words in both languages via their shared phonology. These bilingual phonological neighbourhoods could help facilitate the acquisition of doublets.

2.4.3 Maturational constraints in productive vocabulary growth

Another important finding of this study is that despite bilinguals having larger total vocabulary sizes in comprehension, total vocabulary sizes were similar across monolinguals and bilinguals for production. This finding that monolingual and bilinguals of the same age produce, on average, a similar total number of words (despite bilinguals understanding more words) has interesting implications for the developmental trajectory of word production. These patterns observed in production are likely to be linked to maturational constraints in word production. Toddlers typically understand many more words than they produce. This comprehension-production gap has been observed even in adults when adult participants are taught new words in a novel word learning experiment (Gershkoff-Stowe & Hahn, 2013). Unlike comprehension, which only involves the cognitive process of speech recognition, word production is also dependent on development of motor skills and active rehearsal. Sosa and Stoel-Gammon (2012) found that toddlers' early word productions showed differences in the production accuracy relative to typical adult pronunciations and have variability in pronunciations of the same word. There was more variability found for words that were phonetically more complex and had later-learnt sound combinations. Green, Nip, Maassen, and Van Lieshout (2010) put forward a model of speech development that involves both constraints (milestones in perceptual, cognitive or motor development which limits the learning speed for producing new words) and catalysts (cognitive, neural or environmental supports that help children learn to produce new words). Constraints for speech production by toddlers may include the gradual maturation of articulatory coordination of speech articulators (Green, Moore, Higashikawa, & Steeve, 2000; Iuzzini-Seigel, Hogan, Rong, & Green, 2015). Given that bilingual toddlers are likely to be equally constrained by developmental milestones in fine motor control, it is unsurprising that the total number of words produced does not differ significantly between monolingual and bilingual groups.

In addition to the findings from total vocabulary size, I also found that bilingual toddlers have lower conceptual vocabulary in production than their monolingual peers (significant for the AL-dominant group and marginally significant for the other two groups). Again, this is in contrast to the similar conceptual vocabulary sizes observed between the two groups in comprehension. When considering the possible reasons behind this asymmetry, it is important to note that productive vocabulary as calculated using the Oxford CDI and its adaptations will always be a subset of the child's vocabulary in comprehension. Vocabulary size in production is obtained by counting entries where parents report that their child "understands and says" the word, while vocabulary size in comprehension includes both entries marked as "understands and says" and entries marked as "understands only". It relies on the assumption that a child must understand the word to produce it. As a result, smaller conceptual vocabulary size in production for bilinguals compared to monolinguals despite similar vocabulary sizes in comprehension can be interpreted as bilinguals producing less unique concepts than expected from their vocabulary in comprehension. I propose that this asymmetry could be attributed to a combination of two previously-discussed factors – a preference towards learning translation equivalents and the maturational constraints for productive vocabulary growth. If bilingual toddlers have a preference towards learning the translation equivalents of known words over new concepts, this could explain the slightly smaller conceptual vocabulary found in bilinguals. This is supported by the observation that bilingual toddlers produce a large number of doublets.

2.4.4 Limitations

I acknowledge certain limitations in this study. The CDI is not an exhaustive list of all the words that a child may know, but instead is a subset of commonly-known words aimed to provide an estimate of a child's vocabulary knowledge compared to their peers. Our AL CDIs were adapted from the Oxford CDI, which was normed using data from monolingual British toddlers. As such, the subset of words in the Oxford CDIs (and subsequently our AL CDIs) may be biased towards concepts that are familiar to the UK English-speaking community. While the toddlers in the bilingual sample were also growing up in the UK, there may be certain concepts less common in their home environment due to cultural differences. I attempted to reduce this bias by using only the subset of concepts that was common across all our CDIs after appropriate substitutions were made by native speakers of those languages, but I acknowledge that words common to the UK English-speaking community may still have received greater weight in the calculation of conceptual vocabulary. Further research on bilingual communities would benefit from questionnaires and tasks specifically designed to reflect the experiences of bilingual communities.

There may also be limitations in the use of CDIs to measure vocabulary size of bilingual toddlers in their less-dominant language. In the present study, several parents of bilingual toddlers indicated uncertainty in answering the CDI in their non-native language. Vagh et al. (2009) observed a similar issue, with 16 parents of 118 opting out of reporting their child's English vocabulary due to lack of confidence. This opt-out rate of approximately 10% is similar to the rate observed in the present study. The 42 families who explicitly expressed uncertainty in their reporting accuracy were excluded from the analyses. I also required at least one parent to be a native speaker of the AL, and at least one parent to have fluent English proficiency (operationalised as self-rated proficiency of at least 7 out of 10). Through these criteria, I aimed to reduce the variability in parents' reporting accuracy as a result of low proficiency in one of the target languages.

2.5 Conclusion

In this chapter, I presented analyses on the vocabulary size trajectories of bilingual toddlers relative to monolingual toddlers in English vocabulary, conceptual vocabulary and total vocabulary. Observed differences between groups in English vocabulary size highlights the effect of language exposure on vocabulary growth in a single language. In contrast, comparable conceptual vocabulary sizes in comprehension suggests that bilinguals and monolinguals acquire new concepts at similar rates. Bilinguals also have high total vocabulary sizes, with many words being doublets. The high proportion of doublets in bilinguals' vocabularies points towards a preference towards learning translation equivalents of known words. I also discussed about maturational constraints in the development of productive vocabulary, linking it to comparable total vocabulary sizes in bilingual and monolingual toddlers. The patterns of bilingual vocabulary growth observed in this study highlight both similarities and differences between bilinguals and monolinguals. Both groups show similar trajectories for learning new concepts and development of production. However, a unique aspect of bilinguals' language learning is the availability of cross-linguistic semantic and phonological overlap.

The rest of this thesis will explore the semantic and phonological connections between words in toddlers' vocabularies. In Chapter 3, I will evaluate two algorithms for calculating phonological similarity between words. Following that, the possible effects of cognates on vocabulary trajectories will be explored in-depth in Chapter 5.

Chapter 3

Evaluating two algorithms for cognate classification

3.1 Background

Before I delve into investigating the possible effects of cognates on vocabulary learning trajectories, I need to address the methodological issue of how to define a cognate. Cognates can be similar in phonology, orthography, or both. Some cognates are completely identical in form but oftentimes the definition of cognates also includes word pairs that have high but incomplete overlap. This leaves a degree of uncertainty – how do we define word similarity and what cut-off do we use to classify word pairs as cognates or non-cognates? This is an important methodological question, as standardising our definition of cognates will help the replicability of research on the acquisition, processing and production of cognates.

To further complicate the situation, a cognate may have stronger orthographical overlap than phonological overlap, for example English "chocolate" (/tʃɒkəlɪt/) and Spanish "chocolate" (/tʃokolate/). A cognate could also have stronger phonological overlap than orthographical overlap, for example English "shoe" (/ʃuː/) and German "Schuh" (/ʃuː/). As my thesis focuses on infant bilinguals who have not yet learnt to read, I will discuss cognateness only in relation to phonological similarity, not orthographic similarity nor historical common etymology.

3.1.1 Speaker judgements of phonological similarity

There have been various ways of defining cognates in the literature. One method, commonly used by researchers who primarily work with one language pair, requires native speakers of the language pair of interest (often with linguistic training) to score translation equivalent pairs as cognates or non-cognates. When more than one scorer is employed, the inter-rater reliability of scores for each word pair would then be compared to judge the utility of the cognate classifications. This method has the advantage of drawing upon the intuitions of native speakers. However, the threshold of classifying non-identical cognates can be subjective and may be prone to researcher bias. Additionally, in cases where several different languages are being studied in parallel, it can be difficult to ensure consistency between ratings for each language pair.

Another method that relies on speaker judgements is a translation elicitation task. In contrast to native speaker judgements, translation elicitation tasks use monolingual participants. Given no prior knowledge of the test language, participants are expected to rely on phonological similarity between the presented word and words in their native language as the primary answering strategy. Under this method, words that are correctly translated by a high number of participants are classified as cognates (Kroll et al., 2002). This method for defining cognates is informative in that it reflects real-time processing of auditorily-presented words and their links to a lexicon in a different language. It naturally captures multiple facets influencing cognate recognition – foremost the phonological similarity between two words, but also other factors such as lexical frequency and neighbourhood density, which are known factors that affect word recognition and retrieval. However, this method is time-consuming and resource-heavy as it requires behavioural data to be collected for each word and each language pair.

3.1.2 Algorithms for calculating phonological similarity

Alternatively, there are computed measures that rely on algorithms applied on phonological transcriptions. These algorithms provide the means to apply a standardised and objective method to quantify phonological similarity. Levenshtein distance (Levenshtein, 1966) is a widely-used metric for calculating similarity between words. It calculates word distance using a minimal edit distance calculation that takes into account 3 action types – substitutions, insertions and deletions. The calculation matches identical phonemes wherever possible, accounting for order in sequence. It then calculates the total number of actions needed to change one word to the other, adding a score of 1 for each action. A Levenshtein distance score of 0 means that the words are identical, and the larger the score, the more dissimilar the words are. This can optionally be standardised by dividing by word length, or converted into a similarity score through subtraction.

However, while Levenshtein distance is widely used for calculating orthographic similarity, it has limitations for phonological similarity, particularly in its definition of phoneme mismatches. The key limitation of Levenshtein distance for calculating phonological similarity is that it treats each change equally, without taking into consideration the relative size of the change. An [o] replaced by an [u] is classified as a substitution, and an [o] replaced by a [k] is classified as a substitution of equal weight. Perceptually, however, an [o]-[u] change is likely to be treated as a smaller change than an [o]-[k] change, given that in the former case the two vowels only differ on a single element of vowel height, while the latter replaces a vowel with a consonant. There is experimental evidence suggesting the degree of similarity between two phonemes is not discrete but instead continuous. Findings of unprompted mispronunciation restorations in a shadowing task suggest that mispronunciations in vowels (change in backness) and stress patterns are easily missed, while consonant voicing changes are more obvious (Bond & Small, 1983). Consonant mispronunciations that differ only on 1 feature (as compared to 3) are both easier to restore in a shadowing task and more likely to be missed in a detection task (Marslen-Wilson & Welsh, 1978). Additionally, participants are better at detecting mispronunciations in place of articulation than voicing, and voicing changes are also more obvious when occurring in stops than fricatives (Cole et al., 1978). This is perhaps the biggest challenge in defining phonological similarity between languages. Even if a word pair has shared etymology, historical language change often results in similar but non-identical representations (e.g. [æ] in English “cat” and [a] in Spanish “gato”). This makes Levenshtein distance, which treats all mismatches as equal, limited in its utility for judging phonological similarity. The resulting score from Levenshtein distance may underestimate phonological similarity between words, which can be problematic for cognate identification. In response to this problem, various researchers have proposed alternative algorithms specifically targeted to calculate phonological similarity. These algorithms are commonly used in dialectology. Kondrak (2003) put forward ALINE, an algorithm that aligns non-identical phonemes and gives weights to substitutions according to feature similarity. Through an alignment algorithm that matches consonants to consonants and vowels to vowels where possible, ALINE is able to align phonemes in one word with its closest equivalent in the other word, and subsequently compare phoneme changes with a gradient approach. ALINE gives a score to phoneme substitutions on a continuous dimension using phoneme feature distance, additionally giving weights to different feature change types. This produces a continuous score of how close two phonemes are, which models the way some phonemes are perceptually more similar than others. The 12 phoneme features used for calculating phonetic similarity (with the default salience of the feature in the algorithm in brackets) are manner (50), place (40), voice (10), lateral (10), nasal (10), retroflex (10), syllabic (5), aspirated (5), high (5), back (5), round (5) and long (1). The original ALINE scoring system produced a score that ranged from 0 to Infinity, with strong effects of word length. Larger scores represented higher phonological similarity. Downey, Hallmark, Cox, Norquest, and Lansing (2008) later expanded the ALINE algorithm to produce standardised scores normalised for word length, ranging from 0 to 1.

3.1.3 The Present Study

In this chapter, I evaluate Levenshtein distance (Levenshtein, 1966) and ALINE (Kondrak, 2003; Downey et al., 2008) for calculating phonological similarity, comparing them against the behavioural responses from a translation elicitation task.

I used a translation elicitation task as behavioural validation for the two measures. Monolingual participants provided best-guess matches to native words from presented foreign words. In the absence of knowledge of the language, participants should only be able to guess correctly if there was clear phonological similarity between the presented word and the correct translation. The percentage of participants who give a particular response for a given presented word therefore can be used as a measure for the degree of phonological similarity between the two words. I decided to utilise monolingual judgements to determine the cutoff for cognateness because monolingual listeners have no prior knowledge of the presented language, similar to infant learners just starting to learn their languages. Mimicking the input conditions of infant listeners, participants were not shown the orthographical form of the word, only hearing the presented words auditorily. This removed the availability of orthographical similarity as a cue for participants' answers, making them rely on phonological matching. Audio stimuli were recorded with an infant-directed speech register, which has been associated with exaggerated prosody. I expect that monolinguals will be sensitive to the most salient similarities and largely ignore more subtle overlap.

3.2 Methods

3.2.1 Translation elicitation task

Monolingual participants aged between 18 and 40-years-old were asked to guess the translation of Spanish or Catalan words in their native language. There were three groups: (1) English participants presented with Spanish words (N = 31, 103 trials each); (2) English participants presented with Catalan words (N = 33, 86 trials each); and (3) Spanish participants presented with Catalan words (N = 31, 86 trials each). Spanish and Catalan are very similar languages that share many cognates. In contrast, English shares only a modest number of cognates to Spanish and to Catalan. I therefore expected that Spanish participants listening to Catalan words would perform much better than both groups of English participants. The two English groups should perform similarly. English participants had no prior experience with learning Spanish or other close Romance languages. Spanish participants had no prior experience with Catalan. The task was written in PsychoPy (Peirce, 2007) and hosted online on Pavlovia. English participants were recruited via Prolific (*Prolific*, 2014) and paid £2.50. Spanish participants were recruited from universities in Spain and paid €3. Participants were requested to use headphones in a quiet room and to type their answers using a PC keyboard. On each trial, one word would be presented to participants in auditory format only. The words were recorded with an infant-directed register spoken by a female bilingual native speaker of Catalan and Spanish. Words included in this task were stimuli previously used in word recognition studies with bilingual infants conducted in the lab. These words were concrete nouns with high imageability and high frequency in child-directed speech. Participants were required to provide one answer for each trial by typing. I excluded trials where participants had a high likelihood of having prior knowledge of the word. Examples of these trials were Spanish “agua” which is known by many English speakers to mean “water”. After exclusions, I had 94 trials for the Spanish-English group, 79 trials for the Catalan-English group and 85 trials for the Catalan-Spanish group. The subsets of included words and excluded words are listed in Appendix D.

To define participant accuracy, I used Damerau-Levenshtein distance to compare the orthographic form of the correct answer with participants’ typed answers for each trial. Answers were defined as correct if they match the target answer’s orthography (Damerau-Levenshtein distance = 0). I also accepted typos with 1 edit (substitution, insertion, subtraction or transposition) as correct answers. Typos were defined as having Damerau Levenshtein distance of 1 from the correct answer and not being a legitimate word in the dictionary. For example, “bwol” was accepted as a typo for “bowl”, but “owl” was not accepted as a typo for “bowl” because “owl” is a real word. Following this procedure, each trial for each participant was coded with a binary score of correct or incorrect. The percentage of participants who provided the correct answer for a given trial was calculated by dividing the number of correct answers by the number of participants. Percentages were calculated separately for each trial for each of the three participant groups. English-Spanish trials had a mean participant accuracy of 12.3% (SD = 24.6%); English-Catalan trials had a mean accuracy of 15.9% (SD = 28.5%); Spanish-Catalan trials had a mean accuracy of 46.8% (SD = 41.4%).

3.2.2 Phonological transcriptions

I used the phonological transcriptions of all presented words and their target translations in English and Spanish to generate numerical scores of string similarity using the below computed measures. English words were transcribed according to British English pronunciations, and Spanish transcriptions were transcribed according to European Spanish pronunciations. Stress markers were not included in the transcriptions. Long vowels were duplicated (e.g. “bee” was transcribed as [bii] instead of [bi:]).

3.2.3 Levenshtein similarity

String similarity using the Levenshtein distance method was calculated using the `stringsim` function (method = “lv”) from the `stringdist` package (van der Loo, 2014) in R (R Core Team, 2013). This function calculates the smallest number of insertions, deletions and substitutions needed to change one string into another, generating a distance score. The distance score is then divided by the maximum possible distance to standardise it, and then the result is subtracted from 1 to obtain a similarity score between 0 (no overlapping phonemes) and 1 (identical). The Levenshtein similarity scores between each presented word and its correct translation can be found in Appendix D.

3.2.4 ALINE similarity

String similarity using the ALINE method was calculated using the `alineR` package (Downey, Sun, & Norquest, 2017) in R. The `alineR` package calculates Kondrak's (2003) phoneme-weighted string distance then normalises the distance by word length using Downey et al.'s (2008) algorithm. The result is subtracted from 1 to obtain a similarity score between 0 (maximally dissimilar) and 1 (identical). The ALINE similarity scores between each presented word and its correct translation can also be found in Appendix D.

3.2.5 Statistical tests

I used two metrics to judge the accuracy of the computed measures for cognate identification. Firstly, the numerical scores representing string similarity generated using the Levenshtein and ALINE measures were correlated against the percentage of participants who provided the correct answer for a given presented word. Correlations (Pearson's r) were calculated using the `cor.test` function (R Core Team, 2013) in R.

Secondly, I compared the classification of words into cognates and non-cognates using these computed measures. I used signal detection theory to quantify the ability of Levenshtein similarity and ALINE similarity scores respectively for detecting cognate classifications. Signal detection theory tests the ability of a continuous measure to differentiate between two categories. In the case of this study, the two categories are cognates and non-cognates. Cognate classifications using the translation elicitation task were used as the gold standard for the tests reported in this chapter. Trials where 30% or more participants provided the correct answer were considered cognate trials, while those less than 30% were considered non-cognates. The cutoff of 30% is an arbitrary one. I expect word pairs with high phonological similarity to be considered cognates, and word pairs with low phonological similarity be considered non-cognates. The ability of Levenshtein similarity and ALINE similarity to accurately classify word pairs as cognates or non-cognates would reflect their accuracy as a measure of the phonological similarity perceived by human participants.

Receiver operating characteristic (ROC) curves provide a method to summarise all possible outcomes of the binary classification tool under different thresholds. An ROC curve is created by plotting the measure's sensitivity against its specificity for the various possible thresholds. Trials classified as cognates by both the behavioural task and computed measure were considered True Positives (TP). Trials classified as non-cognates by both the behavioural task and computed measure were considered True Negatives (TN). Trials classified as cognates by the behavioural task but as non-cognates by the computed measure were considered False Negatives (FN). Finally, trials classified as non-cognates by the behavioural task but as cognates by the computed measure were considered False Positives (FP). Using these values, the classification accuracy (Equation 3.1), sensitivity (Equation 3.2) and specificity (Equation 3.3) of each computed measure were calculated. Accuracy represents the proportion of trials that the classification from the evaluated algorithm is the same as the classification from the translation elicitation task. Sensitivity, also known as True Positive Rate, represents the proportion of trials that are classified as cognates according to the translation elicitation task that are also classified as cognates by the algorithm being evaluated. Specificity, also known as True Negative Rate, represents the proportion of trials that are classified as non-cognates according to the translation elicitation task that are also classified as non-cognates by the algorithm being evaluated.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{3.1}$$

$$Sensitivity = \frac{TP}{(TP + FN)} \tag{3.2}$$

$$Specificity = \frac{TN}{(TN + FP)} \tag{3.3}$$

ROC curves also allow us to identify the threshold where the tested measure's classification performance is optimised. This is useful for extending the classifier to new items. Youden's index (Equation 3.4) was used to determine the cut-off for cognate classification using Levenshtein similarity and ALINE. The cut-off corresponded to the point when Youden's index was highest, which maximises the sensitivity and specificity of the measure. The function `coords` from the R package pROC (Robin et al., 2011) was used for calculating Youden's index and identifying the maximum value.

$$Youden's\ index = Sensitivity + Specificity - 1 \tag{3.4}$$

Finally, I compared the classification performance of Levenshtein similarity and ALINE similarity against each other using the Area-under-curve (AUC) of their ROC curves. A higher AUC indicates that the measure is better at predicting cognate classifications from the translation elicitation task. AUC was calculated using the `auc` function from the `pROC` package.

3.3 Results

Using the specified cut-off of at least 30% correct answers, trials in the translation elicitation task were classified for the three groups as follows:

1. English-Spanish: 14 cognates and 80 non-cognates
2. English-Catalan: 13 cognates and 66 non-cognates
3. Spanish-Catalan: 45 cognates and 40 non-cognates

This resulted in a total of 72 trials classified as cognates and 186 trials classified as non-cognates.

3.3.1 Performance of Levenshtein similarity

English-Spanish word pairs had a mean Levenshtein similarity score of 0.121 (SD = 0.169); English-Catalan word pairs had a mean of 0.155 (SD = 0.186); Spanish-Catalan word pairs had a mean of 0.388 (SD = 0.275).

Levenshtein similarity scores had a correlation of .782 with the percentage of participants who provided the correct answer in the translation task (Figure 3.1). The best performance for Levenshtein similarity according to the Youden method had a threshold of 0.310, with a corresponding specificity of 0.889 and sensitivity of 0.898. Translation equivalents pairs with Levenshtein similarity scores greater than 0.310 were considered cognates, while those with score of 0.310 or less were considered non-cognates. As Levenshtein similarity only counts identical phonemes, scores were low for most words in the stimuli set. The optimal cut-off was therefore relatively low to match the distribution of scores. Using this cut-off, there were 83 pairs classified as cognates and 175 pairs as non-cognates. English-Spanish trials had 15 cognates and 79 non-cognates; English-Catalan trials had 20 cognates and 59 non-cognates; Spanish-Catalan trials had 48 cognates and 37 non-cognates. For cognate classification, Levenshtein similarity had an overall accuracy of .895. Distribution of true positives, true negatives, false positives and false negatives are shown in Table 3.1. The high sensitivity of the measure means that classifications using Levenshtein similarity means that almost 90% of the trials classified as cognates according to behavioural results were correctly classified as cognates according to Levenshtein similarity. The high specificity of the measure also indicates that 89% of trials classified as non-cognates by the behavioural measure were correctly classified as non-cognates according to Levenshtein similarity. Cognate classification using Levenshtein similarity had overall high accuracy. This pattern of high sensitivity, specificity and accuracy was seen consistent across the three language pairs tested in the translation elicitation task. The breakdown of accuracy, specificity and sensitivity across the three language pairs is shown in Table 3.2.

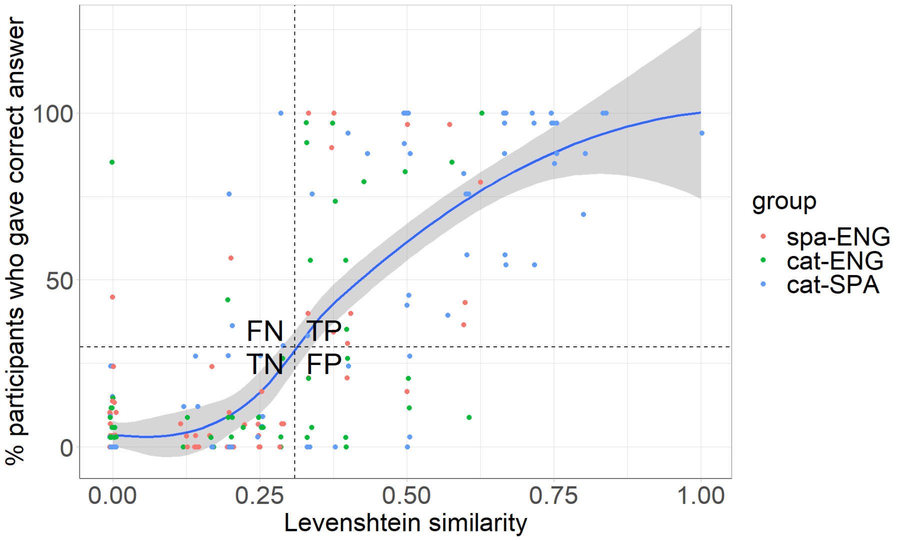

Figure 3.1: Scatterplot of Levenshtein similarity against percentage of participants who gave correct answer, with dotted lines indicating cognate classification cut-offs points and quarters marked for TP, TN, FP, FN

Table 3.1: Distribution of true positives, true negatives, false positives and false negatives when comparing Levenshtein similarity (Lev.) cognate classification against behavioural classification (beh.).

	Lev. cognate	Lev. non-cognate
beh. cognate	64	8
beh. non-cognate	19	167

Table 3.2: Breakdown of accuracy, specificity and sensitivity of Levenshtein similarity for English-Spanish, English-Catalan and Spanish-Catalan pairs.

language pair	sensitivity	specificity	accuracy
English-Spanish	0.857	0.962	0.947
English-Catalan	0.846	0.864	0.861
Spanish-Catalan	0.911	0.825	0.871

3.3.2 Performance of ALINE similarity

English-Spanish word pairs had a mean ALINE score of 0.535 (SD = 0.202); English-Catalan word pairs had a mean of 0.581 (SD = 0.204); Spanish-Catalan word pairs had a mean of 0.737 (SD = 0.211).

ALINE scores had a correlation of $.731$ with the percentage of participants who provided the correct answer in the translation task (Figure 3.2). The best performance for ALINE according to the Youden method had a threshold of 0.797, with a corresponding specificity of 0.901 and sensitivity of 0.952. Translation equivalents pairs with ALINE similarity scores greater than 0.797 were considered cognates, while those with score of 0.797 or less were considered non-cognates. In comparison to Levenshtein similarity, ALINE scores were generally high. This is because ALINE assigns scores based on overlap in phoneme features. Phonemes can vary in the degree of overlap, with 0 meaning maximally dissimilar and 1 being an identical match. Many phoneme pairs would share a small number of features, resulting in a high word similarity score. The optimal cut-off was therefore comparatively high for ALINE to match the distribution of scores. Using this cut-off, there were 74 pairs classified as cognates and 184 pairs as non-cognates. English-Spanish trials had 14 cognates and 80 non-cognates; English-Catalan trials had 17 cognates and 62 non-cognates; Spanish-Catalan trials had 43 cognates and 42 non-cognates. For cognate classification, ALINE had an accuracy of .938. Distribution of true positives, true negatives, false positives and false negatives are shown in Table 3.3. ALINE similarity showed better performance than Levenshtein similarity across all three evaluations of accuracy, sensitivity and specificity. The high sensitivity of the measure means that classifications using ALINE means that more than 90% of the trials classified as cognates according to behavioural results were correctly classified as cognates according to ALINE similarity. Additionally, the high specificity of the measure indicates that 95% of trials classified as non-cognates by the behavioural measure were correctly classified as non-cognates according to ALINE similarity. Cognate classification using ALINE similarity had overall high accuracy. This pattern of high sensitivity, specificity and accuracy was seen across the three language pairs except for sensitivity for English-Spanish word pairs. The breakdown of accuracy, specificity and sensitivity across the three language pairs is shown in Table 3.4. Sensitivity for English-Spanish trials was lower than the other scores, suggesting that there were comparatively more trials misclassified as non-cognates despite being classified as cognates by the behavioural measure. Nevertheless, the sensitivity was still moderately high with 79% of cognate pairs being classified correctly as cognates, and overall accuracy was high with 94% correct classifications.

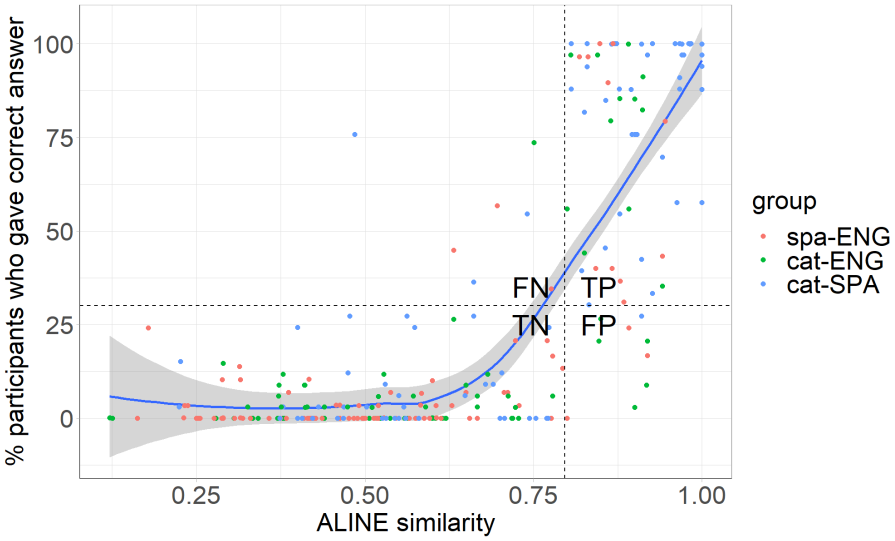

Figure 3.2: Scatterplot of ALINE similarity against percentage of participants who gave correct answer, with dotted lines indicating cognate classification cut-offs points and quarters marked for TP, TN, FP, FN

Table 3.3: Distribution of true positives, true negatives, false positives and false negatives when comparing ALINE cognate classification against behavioural classification (beh.).

	ALINE cognate	ALINE non-cognate
beh. cognate	65	7
beh. non-cognate	9	177

Table 3.4: Breakdown of accuracy, specificity and sensitivity of ALINE for English-Spanish, English-Catalan and Spanish-Catalan pairs.

language pair	sensitivity	specificity	accuracy
English-Spanish	0.786	0.962	0.936
English-Catalan	0.923	0.924	0.924
Spanish-Catalan	0.933	0.975	0.953

3.3.3 Comparing Levenshtein and ALINE

Levenshtein similarity scores had $.829$ correlation with ALINE scores, indicating that they had high agreement on scoring the phonological similarity between word pairs.

I then compared the ROC curves of these two measures to evaluate their classification performance. As previously mentioned, ROC curves plot the measure's sensitivity (true positive rate) against its specificity (true negative rate). The ROC curves for the two measures can be seen in Figure 3.3. Sensitivity is plotted on the y-axis, ranging from 0 to 1. Specificity is plotted on the x-axis in reverse order, from 1 to 0. It is also common in the literature to plot $1-$Specificity on the x-axis; the method used in Figure 3.3 is simply a stylistic difference and presents the same information as plots with $1-$Specificity.

The overall performance of a measure can be quantified using the Area-Under-the-Curve (AUC) of its ROC curve. An AUC of 1 means that the measure is perfectly accurate at discriminating between the two target categories. An AUC of 0 means that the measure is perfectly inaccurate; that is to say that it incorrectly classifies all Category 1 members as belonging to Category 2, and all Category 2 members as belonging to Category 1. An AUC of 0.5 means that has no discrimination ability, with category classification mostly random. An AUC of 0.8 to 0.9 is considered to mean very good discrimination, and over 0.9 to be outstanding (Hosmer Jr, Lemeshow, & Sturdivant, 2013).

AUC of the ROC curve for Levenshtein similarity predicting behavioural cognate classification was 0.941. AUC of the ROC curve for ALINE similarity predicting behavioural cognate classification was 0.961. As such, ALINE is slightly better than Levenshtein similarity at classifying cognates, but the difference is small.

3.4 Discussion

The cognate classification accuracy of both Levenshtein similarity and ALINE were high, at 89.5% and 93.8% respectively. Both measures also had high positive correlations with the percentage of participants who provided the correct answer for each trial. This supports the utility of these measures of string similarity in predicting monolingual participants' success at matching an unfamiliar foreign word to a word in their native language using only phonological information.

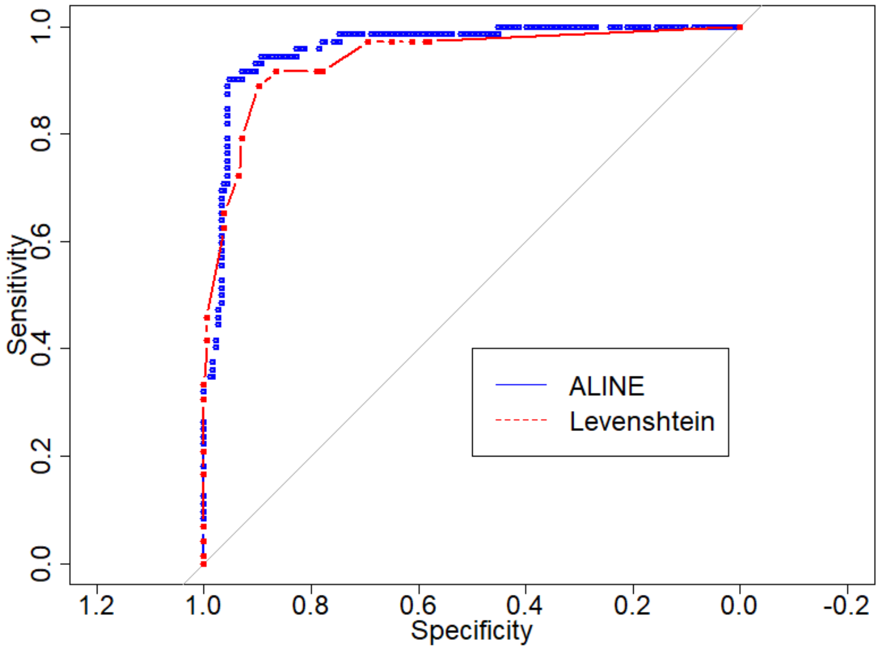

Figure 3.3: ROC curves of ALINE and Levenshtein for predicting cognateness.

3.4.1 Discrete vs gradient phoneme changes

It is interesting that despite providing a rough comparison of phonological overlap (only counting identical matches and not using any fine-grained calculations of phoneme closeness), the accuracy of the classification criteria from Levenshtein similarity was very similar to that of ALINE. This suggests that while listeners may be sensitive to some gradient changes in phoneme closeness, their responses are largely influenced by identical phonemes. This suggests that the proportion of identical matches may be a good predictor of participants' success at cross-linguistic lexical matching. The high correlation ($.823$) between ALINE scores and Levenshtein scores also suggests that inclusion of fine-grained phoneme closeness calculation does not drastically change the relative similarity of different word pairs.

However, there are some word pairs that Levenshtein underestimates the similarity of, listed in Table 3.5. These pairs are characterised by a high proportion of non-identical but close phonemes. This, alongside the slightly higher accuracy ALINE has compared to Levenshtein for cognate classification, suggests that fine-grained phoneme calculations does improve algorithms for phonological similarity relative to discrete phoneme overlap.

Table 3.5: List of trials where Levenshtein similarity classifies the pair as a non-cognate but ALINE classifies as cognate and there is high participant accuracy. Words are listed with their IPA transcriptions in brackets.

presented	correct	group	lv	aline	%acc
mirall (/miɾaʎ/)	mirror (/mɪrə/)	C-E	0.20	0.83	44.1
girafa (/ʒiɾafə/)	giraffe (/ʤɪrɑɑf/)	C-E	0	0.90	85.3
esponja (/əspɔnʒə/)	esponja (/espoŋxa/)	C-S	0.29	0.83	30.3
formiga (/furmigə/)	hormiga (/oɾmiɣa/)	C-S	0.29	0.82	100

Note. lv = Levenshtein similarity; %acc = % participants who gave correct answer; S-E = Spanish stimuli – English participants; C-E = Catalan stimuli – English participants; C-S = Catalan stimuli – Spanish participants

Conversely, the word pairs in the stimuli that ALINE classified as non-cognates but Levenshtein classified as cognates are listed in Table 3.6. These trials feature ALINE scores that are high (exceeding 0.7) but fall below the threshold of 0.797 assigned according to the highest value of Youden's index. These word pairs have several identical phonemes but are penalised by having clear dissimilarity in other parts of the word. The ratio of matching phonemes and non-matching phonemes resulted in a moderately high similarity score for both Levenshtein and ALINE near the threshold.

Table 3.6: List of trials where ALINE classifies the pair as a non-cognate but Levenshtein classifies as cognate and there is high participant accuracy. Words are listed with their IPA transcriptions in brackets.

presented	correct	group	lv	aline	%acc
ensalada (/ensalada/)	salad (/sæləd/)	S-E	0.38	0.78	34.5
tomaquet (/tumakət/)	tomato (/təmɑɑtəʊ/)	C-E	0.38	0.71	73.5
pa (/pa/)	pan (/pan/)	C-S	0.67	0.74	54.5

Note. lv = Levenshtein similarity; %acc = % participants who gave correct answer; S-E = Spanish stimuli – English participants; C-E = Catalan stimuli – English participants; C-S = Catalan stimuli – Spanish participants

3.4.2 Limitation in phonological transcriptions

There were a small number of trials that both ALINE and Levenshtein erroneously classified as non-cognates despite there being high participant accuracy. Table 3.7 lists the pairs that are considered non-cognates by both Levenshtein and ALINE cognate classification, but have more than 30% participant accuracy. These word pairs feature many close phonemes and few (if any) identical phonemes. This resulted in a very low Levenshtein similarity score and a moderate ALINE score that did not reach the cognate threshold. However, despite these differences in phonemes, many participants were able to identify the correct translation. This suggests that the phonological similarity between some word pairs may not be aptly captured by the string similarity between the phonological transcriptions of these words. Phonological transcriptions are a simplification of the spoken word signal that listeners hear, and may not capture different nuances that may affect listeners' perception. Additionally, a key limitation of both Levenshtein and ALINE is that they do not take into account the stress patterns of the two words, which may result in underestimation of phonological similarity. Despite having high accuracy, judgements of cognates using string similarity algorithms should still be used with caution.

Table 3.7: List of trials where both Levenshtein similarity and ALINE both classify the pair as a non-cognate but there is high participant accuracy.

presented	correct	group	lv	aline	%acc
naranja (/naɾaŋxa/)	orange (/ɒrɪnʤ/)	S-E	0	0.631	44.8
tigre (/tiɣɾe/)	tiger (/taɪgə/)	S-E	0.200	0.696	56.7
arbre (/aßɾə)	arbol (/arbol)	C-S	0.200	0.485	75.8
orella (/uɾɛʎə/)	oreja (/oɾexa/)	C-S	0.200	0.661	36.4

Note. lv = Levenshtein similarity; %acc = % participants who gave correct answer; S-E = Spanish stimuli – English participants; C-E = Catalan stimuli – English participants; C-S = Catalan stimuli – Spanish participants

3.4.3 Competition effects

Interestingly, looking at false positives of the algorithms relative to the behavioural task, there is a subset of trials where both Levenshtein similarity and ALINE consider the presented word to have high phonological similarity with its translation but there is low participant accuracy for providing the correct translation. These trials are listed in Table 3.8.

These trials reflect an interesting feature of lexical retrieval in spoken word recognition. In the translation elicitation task, participants are prompted to retrieve a matching word from their lexicon without explicit restrictions. Research into monolingual spoken word recognition has shown that lexical retrieval can be affected by other words in the lexicon. In behavioural studies, when a target word has a high number of competing candidate words, the successful selection of

Table 3.8: List of trials where both Levenshtein similarity and ALINE both classify the pair as a cognate but there is low participant accuracy, along with the word that was given most consistently as an answer for that trial.

presented	correct	group	lv	aline	%acc	top answer	%top
bol	bowl	S-E	0.50	0.92	16.7	ball	76.7
plat	plate	C-E	0.60	0.92	8.8	plait	20.6
bol	bowl	C-E	0.50	0.92	20.6	ball	73.5
flor	flower	C-E	0.40	0.85	26.5	floor	35.3
estruc	ostrich	C-E	0.33	0.90	2.9	strawberry	11.8
gat	cat	C-E	0.33	0.85	20.6	god	23.5
caixa	caja	C-S	0.50	0.91	27.3	casa	69.7

Note. lv = Levenshtein similarity; %acc = % participants who gave correct answer; %top = % participants who gave top answer; S-E = Spanish stimuli – English participants; C-E = Catalan stimuli – English participants; C-S = Catalan stimuli – Spanish participants

the target word is less likely and more effortful compared to the retrieval of targets with few or no competitors (Luce & Pisoni, 1998; Norris, McQueen, & Cutler, 1995; Vitevitch, Stamer, & Sereno, 2008). Several models of spoken word recognition have proposed that when listeners hear a spoken word, they activate several candidate words in the lexicon before finally narrowing the selection down to the most closely matching word. This is a common assumption for the Cohort Model (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987), Shortlist Model (Norris, 1994) and TRACE (McClelland & Elman, 1986). In these models, candidate words compete for activation with the target word, creating lexical interference. Lexical competition exerts strong inhibition effects when a competitor word has high number of matches and few mismatches with the target word (Dufour & Peereman, 2003b). Priming effects are also larger when the target has few competitors (Dufour & Peereman, 2003a).

In the context of the translation elicitation task, I expect that when the presented word has strong phonological similarity to its translation with no strong competitors, participant accuracy should be higher than when the target translation has strong competitors. Particularly, I expect that target translation retrieval would be lowest when there is a competitor word with closer phonological similarity than the correct translation. Taking an example from Table 3.8, the retrieval of the correct translation "bowl" for the Spanish word "bol" is strongly affected by the English competitor word "ball", which is perceived by 77% of participants to be closer to the presented word than "bowl". These cross-linguistic word pairs, where the words have high phonological similarity but different meanings, are commonly referred to as false friends. The interference exerted by the false friend "ball" resulted in comparatively low participant accuracy (17%) despite the strong phonological similarity between "bowl" and "bol".

Lexical interference from false friends does not only affect monolingual listeners with no prior knowledge of the language. Bilingual adult participants have been found to be worse at translating false friends compared to both cognates and non-cognates (Otwinowska & Szewczyk, 2017). The interference effect of false friends can be linked to the non-selective lexical access hypothesis for bilingual word retrieval, which posit that bilinguals activate their lexicon across both languages simultaneously, even when the task only uses one language (Costa et al., 2000). This would result in competitors being activated across both languages. As such, research on the learning and retrieval of cognates should also take into consideration possible effects of lexical competitors in addition to phonological similarity between a cognate pair. An algorithm for cognate classification that accounts for both phonological similarity and lexical competition effects will be explore and evaluated in Chapter 4.

3.5 Conclusion

In this chapter, I evaluated two algorithms for calculating phonological similarity – Levenshtein similarity and ALINE. Both algorithms had high accuracy for modelling the cognate classifications of behavioural responses from a translation elicitation task. ALINE, which includes calculations for gradient phoneme changes, had slightly higher accuracy and sensitivity than Levenshtein, which only counts identical phonemes. However, the difference between the classification accuracy of these two algorithms was smaller than expected, suggesting that monolingual listeners' answers are largely dependent on the presences of identical phonemes between two words.

Due to the better performance of ALINE similarity over Levenshtein similarity, ALINE similarity will be used to classify word pairs into cognates and non-cognates in Chapter 5, separated according to the optimal threshold identified in this chapter.

Chapter 4

An alternative approach to defining cross-linguistic phonological similarity using a model of monolingual speech recognition

4.1 Background

In this chapter, I evaluate the utility of the TRACE model (McClelland & Elman, 1986; Strauss, Harris, & Magnuson, 2007) for modelling monolingual participants' answers in the translation elicitation task described in Chapter 3. In this task, participants listen to unfamiliar foreign words and need to match the foreign input to the closest word in their native language lexicon. Having no prior knowledge of the presented foreign language, participants are left with phonological similarity as their primary source of information. Participant response patterns in translation elicitation tasks have been used as a measure of cognateness, where cognates have a much higher likelihood of being guessed correctly due to the phonological overlap between the presented word and its correct translation. In Chapter 3, I showed that participant responses had high agreement with two algorithms for calculating phonological similarity – Levenshtein distance and ALINE. However, there were also trials that both of these algorithms performed poorly on. A common trait of some of these trials was the existence of a strong competitor that reduced participants' likelihood of choosing the correct translation as their answer. The effect of competitors on cognate classification is not a parameter that is included in either Levenshtein distance or ALINE. This led me to the exploring the possibility of using a computational model of spoken word recognition for modelling participant responses.

Lexical competition effects have formed a core assumption of various computational models of spoken word recognition such as the Cohort Model (Marslen-Wilson & Welsh, 1978), TRACE (McClelland & Elman, 1986) and Shortlist Model (Norris, 1994). These models have typically been used to model recognition of words through matching a presented phonological form to its lexical entry in the listener's existing vocabulary. Recognition success is shown by the correct lexical entry being selected above other words in the lexicon for a given presented phonological form. While these models differ in the algorithms and weights used for calculating the activation of words, they share fundamental similarities. One is the simultaneous activation of multiple candidate words. Hearing the onset phoneme of a input word activates a set of lexical representations in the mental lexicon containing that phoneme. Candidates compete with each other for selection, exerting inhibition of other words. As more of the input word is heard, candidates that match the input gain activation, while candidates that have mismatches reduce in activation. Activation values are calculated for each candidate word until the end of the presentation phase, and the lexical representation with the highest activation (typically the word with the highest number of phoneme matches with the input) would be selected by the model as the final candidate.

Given that activation values are calculated for each candidate word in the lexicon, regardless of whether it fully matched with the input, these models should theoretically be extendable to foreign word perception by monolingual speakers. In the absence of an exact match available in the lexicon, the candidate word that receives the highest activation should be the word that is the closest approximation to the input. For the rest of this chapter, I will explore the TRACE model for modelling participant responses on the translation elicitation task using its Java re-implementation jTRACE (Strauss et al., 2007). The jTRACE re-implementation makes the model easy to run even without advanced programming knowledge. A new lexicon and new phonemes can be easily added to the model. Additionally, . This accessibility is important for this study's long-term purpose to utilise the model as an algorithm for cognate classification in bilingual research.

4.1.1 Parameters of TRACE

TRACE is a computational model that has 3 levels: *features*, *phonemes* and *words*. It is made of processing units that interact with each other both across levels and within levels. Units have *excitatory* connections with units on different levels that are consistent with it. For example, an input of the sound [k] will trigger activation of the features corresponding to it, which in turns activate [k] on the phoneme level and the words that contain [k] on the word level. These connections are bidirectional, so that word level activation can also influence phoneme level activation. Meanwhile, units on the same level that are inconsistent have *inhibitory* connections. The phoneme [k] will inhibit phonemes that are not [k], and words that do not contain [k] will inhibit other words that do not contain [k]. There is no inhibition applied between levels. This combination of excitatory and inhibitory connections generates a predicted activation for each word in the model's lexicon in response to a given input.

Phoneme feature sensitivity

An important element of TRACE (and the re-implementation in jTRACE) is the feature level. TRACE represents phonemes on 7 feature dimensions: Consonantal, Vocalic, Diffuseness, Acuteness, Voicing, Power and Burst. Each feature dimension has 8 units corresponding to it (except for Burst, which has 5 units). Phonemes that are more similar to each other overlap in activation on more units. For example, the voiceless bilabial stop [p] and the voiced bilabial stop [b] differ only on 2 feature dimensions (Voicing and Burst), while [p] and the voiceless alveolar fricative [s] differ on 6 out of 7 dimensions. TRACE takes a given input phoneme and transforms it into features; the activation of feature units then feeds excitatory activation to phonemes that match those features. In this manner, a phoneme which matches on all units will be the strongest activated, but other phonemes with partial matches will also be activated albeit to a lesser degree. This is similar to other computed measures of gradient phonological similarity like ALINE, making it appropriate for my purposes.

Incremental input

TRACE uses time slices to model the incremental nature of spoken language. The presentation of an input string is distributed across these time slices, and the model processes consecutive time slices one at a time in order. Each time slice corresponds to roughly 10 msec of real time. A hypothesis about the most likely lexical item that the input reflects is generated for each time slice, defined by each lexical item's activation value. This hypothesis is affected by previous cycles, thus modelling incremental processing. The effect of phoneme matches and phoneme mismatches therefore manifests in TRACE as increases and decreases in activation as the model unfolds. Matching phonemes increase the cumulative strength of a candidate word's activation through excitatory connections, increasing its chance of being the selected candidate. However, if mismatches occur, the activation will be low or even in negative, therefore decreasing the cumulative activation as a function of inhibitory connections. Each lexical item in the model's lexicon will receive a cumulative activation value that reflects the sum of these excitatory and inhibitory connections across all time slices, as triggered by the input. Accordingly, the more matches and less mismatches a given lexical item has with the presented form, the higher its cumulative activation value would be in TRACE.

Alignment

TRACE allows for lexical entries to be matched to later segments of presented phonological forms, such that a phonological form might be multiple words rather than just one. This feature is important for cognate identification, as there are many ways that cognates could have partial overlap. One word may be embedded in the other, either at the word onset (e.g. Spanish “moto” and English “motorbike”) or later in the word (e.g. Spanish “ensalada” and English “salad”). This feature of TRACE will allow us to test TRACE’s ability to predict participants’ recognition of embedded cognates.

Competition

TRACE models the effect of lexical competition through its within-level inhibitory connections. Words that are activated exert inhibitory effects on other words. Therefore, given the same amount of activation received from the phoneme level, words with more competitors are activated slower as a result of strong within-level inhibition. This parameter of within-level inhibition means that for the same amount of phoneme matches between the input string and a given lexical item, a lexical item with many competitors will receive greater inhibition and therefore achieve lower activation than items with few competitors. Additionally, the more similar the competitors are to the input string, the more inhibition will be applied.

This models the effect of phonological neighbourhood density on word retrieval.

4.1.2 The Present Study

The aim of this chapter was to present and test the utility of the jTRACE model (Strauss et al., 2007) as a test for cognateness. Unlike the algorithms evaluated in Chapter 3, the jTRACE model includes parameters for within-level inhibition designed to mimic the effect of phonological competitors on speech recognition. With this advantage in mind, I aimed to test jTRACE's ability to model participants' responses in a *translation elicitation task*. I evaluated the ability of jTRACE to select the most likely candidate word in response to foreign word input. I compared jTRACE's cognate identification accuracy against the two measures of phonological similarity evaluated in Chapter 3. If successful, jTRACE could offer a computational alternative for calculating cross-linguistic phonological similarity. The graphical interface of jTRACE makes it an accessible option to researchers with or without programming experience. New phonemes and new lexical items can be uploaded to the model via the graphical interface, making the phoneme set and model lexicon fully flexible for adaptations across multiple languages.

4.2 Study 1: Word activation

To test the validity of using jTRACE for modelling cognate identification, I test whether participant responses in the translation elicitation task can be predicted by the relative activation values of the lexical items as modelled by jTRACE's parameters. In the translation elicitation task, participants need to match up the unfamiliar foreign word that they hear with a word in their native language's lexicon. Under the model, an input word triggers excitatory and inhibitory connections that gives activation to candidate words in the mental lexicon. Activation values are predominantly affected by matches and mismatches of the input string to the lexical item. Behaviourally, the final answer given by a participant, whether correct or incorrect, should be the candidate word that is most strongly activated. If there are multiple words in the lexicon with similar levels of phonological similarity with the presented words, we expect to see participant answers spread across multiple lexical items rather than one dominant answer, and for the activation to be similarly spread across multiple candidate words in the model.

4.2.1 Methods

Translation Elicitation Task

Participant data from the translation elicitation task described in Chapter 3 was used as the gold standard for model evaluation. In this task, monolingual participants were presented with unfamiliar foreign words and asked to guess the equivalent translation in their native language. My aim was to evaluate the accuracy of jTRACE for modelling participant responses when matching a foreign word input to a word in their native language lexicon. As the model lexicon tested in this study was an English lexicon, only data from English participants was used for the analysis. This consisted of 31 participants who completed the Spanish version of the task and 33 participants who completed the Catalan version, as reported in

jTRACE

The jTRACE model (Strauss et al., 2007) is a re-implementation of the TRACE model (McClelland & Elman, 1986) in Java. Input words are inserted to the model as a string of computer-readable phoneme transcriptions. The model then sequentially converts each phoneme in the string to features, activating relevant features in the feature level. The 7 feature dimensions are Consonantal, Vocalic, Diffuseness, Acuteness, Voicing, Power and Burst. Each feature dimension has 9 units corresponding to it, for a total of 63 orthogonal units. The units are presented as distributed representations. For any given unit for a given phoneme, the unit is activated if it is relevant to the phoneme, and not activated if it is irrelevant. Activation on each unit ranges from 0 (not activated) to 1 (fully activated). The activation of units in this model is sparse – in each set of 9 units corresponding to a feature, only one unit of the set will have an activation of 1. For some features, for example the Burst feature which is only applicable to stop consonants, there may be no activation for some phonemes. In addition to the activation of 1 for the most relevant unit, some phonemes also have activation values of 0.5, 0.3 or 0.1 distributed over additional adjacent units within the 9 units of the same feature.

The activated features then feeds activation to relevant phonemes, which in turn feeds activation to relevant words from a pre-defined lexicon. Simultaneously, inhibition occurs for non-relevant phonemes and words. This activation and inhibition occurs incrementally over cycles in response to the input.

The first published implementation of the TRACE model in jTRACE by Strauss et al. (2007) used the phoneme set of 10 consonants and 8 vowels specified in McClelland and Elman (1986). Mayor and Plunkett (2014) later expanded the phoneme set to include the full set of phonemes used in British English, adding 12 more consonants and 8 more vowels. The list of phonemes in IPA transcription matched to their jTRACE transcriptions can be found in Appendix E.

The input words were presented words from the translation elicitation task. I inserted input words to the jTRACE model with one silence phoneme preceding and another one following the word, following the convention of Strauss et al. (2007). The subset of trials analysed in this chapter comprised of 40 Catalan-English trials and 26 Spanish-English trials. The list of input words can be found in Appendix E. The subset was necessary because there are currently limited phoneme inventories publically available for jTRACE. I utilised the British English phoneme inventory implemented by Mayor and Plunkett (2014). Unfortunately, this meant that not all the words in the task could be inserted in the model. For the current study, I excluded input words in the translation elicitation task that contain phonemes not currently implemented in the model. This allowed me to perform an initial evaluation of the model's prediction accuracy before undertaking work adding new phonemes to the model. Expanding the model's phoneme inventory to include phonemes from other languages is a point for future work. In addition to the Spanish and Catalan presented words, I also included the English translation for each of these 66 words as input words to evaluate the model's accuracy for selecting words that exist in the model's lexicon.

Due to the constraint of limited phonemes, I only evaluated the model in comparison to English participants' responses. The model's lexicon was built from all answers given by English participants in the translation elicitation task that were single English words. I obtained phonological transcriptions for each word in the lexicon from the Carnegie Mellon University Pronouncing Dictionary. The CMU Pronouncing Dictionary is a open-source database of phonological transcriptions for a large number of English words, transcribed in ARPAbet. For use in the jTRACE interface, each transcription was converted from ARPAbet to the transcription codes used by jTRACE. A limitation of the choice of the CMU Pronunciation Dictionary for the purposes of this study was that the words in the dictionary were transcribed according to North American pronunciations. However, I was unable to identify a suitable alternative for British English pronunciations. Dictionary resources for British English either had a short list that did not cover all the target words, or was not available as a downloadable word list. I ultimately decided that the CMU Pronunciation Dictionary was the best of the available choices. The model lexicon had 1322 words. The list of words with transcriptions and frequency were converted to jTRACE-readable xml format using the `jtracer` (Garcia-Castro, 2021) R package.

I used the default settings of jTRACE for the activation and inhibition applied by lexical frequency and competition for the simulations in this chapter. Resting activation level for features and phonemes prior to input was set at -0.1. Resting activation level for words was -0.01. The maximum possible activation that any unit in the model could achieve was 1, and the minimum activation was -0.3. Activation from input to the feature level had weight of 1. Activation from the feature level to the phoneme level had a weight of 0.02. Activation of the phoneme level to the word level was 0.05. There was top-down activation from word level to phoneme level set at 0.03 but no top-down activation from phoneme level to feature level. Feature-level inhibition was 0.04; phoneme-level inhibition was 0.04; word-level inhibition was 0.03. Feature decay was set at 0.01; phoneme decay was 0.03; word decay was 0.05. The spread of each feature type was set at 6, equal for all features. Lexical frequency was not set as an active parameter in the default model. For more details, see the original paper by Strauss et al. (2007).

The jTRACE model offers multiple options for its data output. Numerical data can be obtained for both activation strength and response probability, with values calculated for each cycle. I selected the maximal alignment output (ad-hoc) option, which generates the maximal value of a word's activation on each cycle. This means that we are able to compare the cumulative activation achieved by different stages of the word's presentation. I extracted the activation of all output words in the model's lexicon after the cycle corresponding to the end of the input word's presentation phase. The input words from the task had lengths ranging from 3 phonemes to 7 phonemes. As each phoneme was presented over 11 cycles, the presentation phase for each word was calculated by the following formula (Equation 4.1):

$$presentation\, phase = 11 \, * \, number\ of\ phonemes \tag{4.1}$$

Levenshtein similarity

Levenshtein similarity scores were calculated using the `stringdist` package (van der Loo, 2014) as in Chapter 3. For direct comparability to the jTRACE output, I used the same subset of words as described in the jTRACE section above. I also used the transcriptions from the CMU Pronouncing Dictionary for English words.

ALINE

ALINE scores were calculated using the `alineR` (Downey et al., 2017) as in Chapter 3. Again, I used the same subset of words as described in the jTRACE section above and the transcriptions from the CMU Pronouncing Dictionary for direct comparability to the jTRACE output.

Statistical tests

As in Chapter 3, I used Receiver Operating Characteristic (ROC) analysis to compare the predictive performance of jTRACE, Levenshtein distance and ALINE relative to participant responses in the translation elicitation task. ROC analysis assesses the overall performance of a diagnostic measure over a range of possible cut-offs. The ROC curve can be characterised using its Area-Under-Curve (AUC), which can vary in range between 0 and 1. An AUC value of 0.5 indicates that the measure has poor discriminatory ability across different cutoffs. Positive values that are significantly different from 0.5 are considered to indicate good performance, with higher values indicating better performance (Mandrekar, 2010). The ROC curve also allows researchers to identify the optimal cutoff threshold for categorising a continuous score into two categories. One common method for identifying a cutoff is using the point on the curve where Youden's index is the highest. The highest value of Youden's index for a measure indicates the cutoff where the measure has the best predictive ability when sensitivity and specificity are weighted equally. I used the `pROC` (Robin et al., 2011) package in R to identify the optimal cutoff using Youden's index.

In addition to the AUC, I compared the specificity, sensitivity and accuracy of the three measures at the optimal cutoffs as judged by Youden's index. The equations for specificity, sensitivity and accuracy were previously described in Chapter 3.

4.2.2 Results

Using the specified cut-off of at least 30% correct answers, trials in the subset were classified for the two English participant groups as follows:

1. English-Spanish: 5 cognates and 21 non-cognates
2. English-Catalan: 9 cognates and 31 non-cognates

This resulted in a total of 14 trials classified as cognates and 52 trials classified as non-cognates.

Performance of jTRACE for English input words

To validate the model, I first tested the model's accuracy for identifying English input words. The difference between the activation of the matched output word and the non-identical word that received the most activation is visualised in Figure 4.1. The jTRACE model successfully selected the correct candidate word in response to English input words in all cases except one.

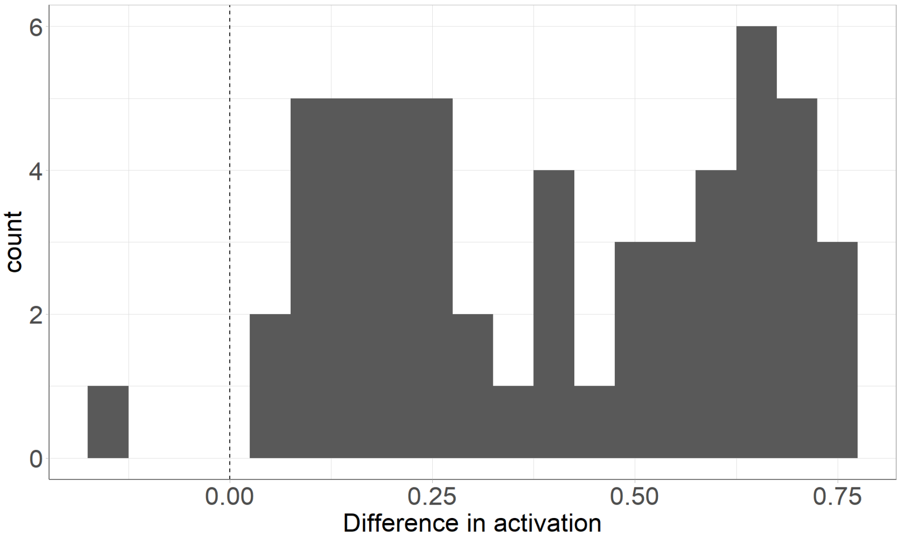

Figure 4.1: Histogram showing the difference in activation between the matched output word and the most highly activated non-identical word. Positive values indicate that the matched word had higher activation than the closest non-identical word.

The word "owl" (aUl) had lower activation than the word "eye" (aI) in response to input of "owl" (aUl). The activation level of "owl" fell further after the end of the presentation phase, suggesting that the word had been inhibited by the erroneously high activation of "eye". As lexical frequency was not an active parameter in the model used for these simulations, the higher lexical frequency of "eye" over "owl" is unlikely to be the cause of this issue. The poor performance for this input word may be related to issues in the phoneme specifications in the model resulting in sub-optimal phoneme-to-word activation, problems with the within-level inhibition strength, or a combination of the two. This will be further explored in the second section of this chapter.

Performance of jTRACE for Spanish/Catalan input words

English-Spanish word pairs had a mean activation score of 0.084 (SD = 0.22) at the end of the presentation phase and English-Catalan word pairs had a mean of 0.057 (SD = 0.20). This meant that jTRACE judged English-Spanish word pairs in the stimuli set to be slightly more similar than English-Catalan word pairs were. Nevertheless, the large standard deviations indicate that both language pairs show variation in their similarity scores between word pairs.

jTRACE activation had a correlation of $.724$ with the percentage of participants who provided the correct answer in the translation task (Figure 4.2). When the correct translation was given as the answer by a high proportion of human participants, jTRACE was moderately successful at also giving a high activation score to the candidate word corresponding to the correct answer. When few participants gave the correct translation as their answer, jTRACE activation for the corresponding candidate word was also low.

I then analysed the model's ROC curve to identify the optimal threshold for classifications of cognateness. The best performance for jTRACE activation according to the Youden method had a threshold of 0.234, with a corresponding specificity of 1 and sensitivity of 0.571. Under this cutoff, translation equivalents pairs with jTRACE activation greater than 0.234 were considered cognates, while those with score of 0.234 or less were considered non-cognates. There were 8 pairs classified as cognates and 58 pairs as non-cognates. English-Spanish trials had 4 cognates and 22 non-cognates. English-Catalan trials had 4 cognates and 36 non-cognates. For cognate classification, jTRACE had an overall accuracy of .909. Distribution of true positives, true negatives, false positives and false negatives are shown in Table 4.1.

Specificity, or the True Negative Rate, represents the proportion of trials that are classified as non-cognates according to the participant behaviour in the translation elicitation task that are also classified as non-cognates by jTRACE. Specificity of 1 means that jTRACE did not incorrectly classify any behavioural non-cognates as cognates. On the other hand, sensitivity, or True Positive Rate, represents the proportion of trials classified as cognates according to the translation elicitation task that are also classified as cognates by jTRACE. In comparison to its perfect specificity for this set of words, jTRACE had a low sensitivity, incorrectly classifying 43% of behavioural cognates as non-cognates.

Table 4.1: Distribution of true positives, true negatives, false positives and false negatives (jTRACE subset) when comparing jTRACE activation (jT.) cognate classification against behavioural classification (beh.).

	beh cognate	beh non-cognate
jT cognate	8	36
jT non-cognate	0	22

An interesting point of note about jTRACE predictions is that the model did

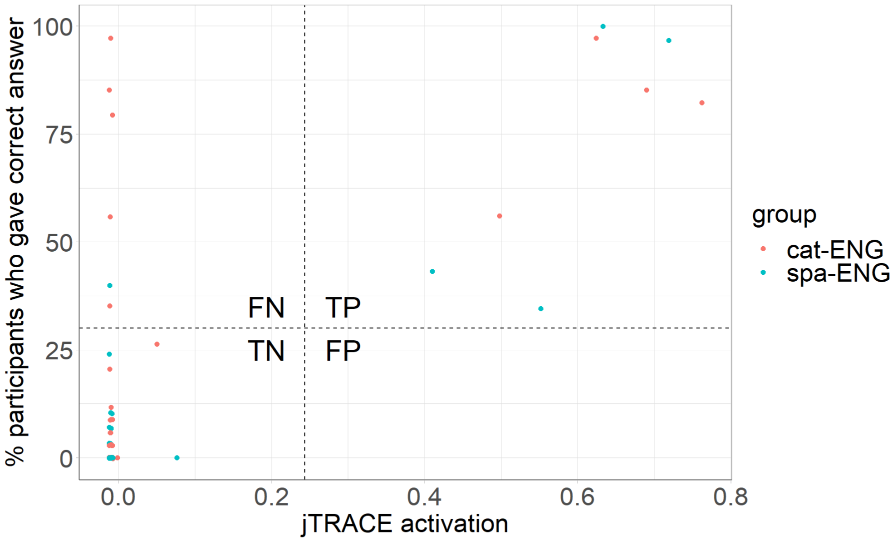

Figure 4.2: Scatterplot of jTRACE activation against percentage of participants who gave the correct answer (jTRACE subset), with dotted lines indicating cognate classification cut-offs points and quarters marked for TP, TN, FP, FN. Each point represents one trial.

not produce any false positives for this set of words. In comparison, there was a high rate of false negatives. Looking at the distribution of jTRACE scores in Figure 4.2, we see that the word pairs that fall in the false negative (FN) quadrant all have scores close to 0, while the word pairs in the true positive (TP) quadrant have much higher scores above 0.4. This pattern shows that jTRACE grossly underestimates the similarity of some word pairs. Table 4.2 lists the words that have high participant accuracy but very low jTRACE activation.

One observation is that words in this list have partial but non-continuous overlap. It is also notable that most of the overlapping phonemes are consonants while the differing phoneme are vowels. This suggests that there may be a consonant bias in participants' lexical retrieval, a feature that is not implemented in jTRACE.

Table 4.2: Trials that have very low jTRACE activation (-0.01) but high participant response rate ($>$ 30%).

word	response	activation	% participant
tren (trDn)	train (trEIn)	-0.01	35.29
chaqueta (tSakEta)	jacket (dMak^t)	-0.01	40
tigre (tigr@)	tiger (taIg@)	-0.01	55.88
esponja (@spLnM@)	sponge (sp^ndM)	-0.01	79.41
girafa (Miraf@)	giraffe (dM@af)	-0.01	85.29
xocolata (Sukulat@)	chocolate (tSLkl^t)	-0.01	97.06

In some cases, there may be additional effects of lexical stress making certain parts of the more salient to listeners than others. Participants may be more inclined to choose an answer that overlaps in the stressed later syllable over one that overlaps at the unstressed word onset. Additionally, participants may also be sensitive to the number of syllables in the word.

In jTRACE, candidate words are penalised with within-level inhibitory effects from competitors with more activation. In these cases, it is likely that competitor words that have more continuous overlap early in the word gained activation faster than the correct answer with partial non-continuous overlap. While the correct answer received some activation from the shared phonemes, the inhibitory effects from competitors were stronger than the activation gained, suppressing the final activation of the correct answer back down to baseline levels.

Performance of Levenshtein similarity

English-Spanish word pairs had a mean Levenshtein similarity score of 0.25 (SD = 0.21) and English-Catalan word pairs had a mean of 0.18 (SD = 0.21).

Levenshtein similarity scores had a correlation of .727 with the percentage of participants who provided the correct answer in the translation task (Figure 4.3). The best performance for Levenshtein similarity according to the Youden method had a threshold of 0.155, with a corresponding specificity of 0.673 and sensitivity of 1. Translation equivalents pairs with Levenshtein similarity scores greater than 0.155 were considered cognates, while those with score of 0.155 or less were considered non-cognates. Using this cut-off, there were 31 pairs classified as cognates and 35 pairs as non-cognates. English-Spanish trials had 12 cognates and 14 non-cognates. English-Catalan trials had 19 cognates and 21 non-cognates. For this subset of words, Levenshtein similarity classified more words as cognates than jTRACE did. Levenshtein similarity only counts identical phoneme matches, disregarding any non-exact matches and effects of lexical competition. In contrast to jTRACE, which had perfect specificity and medium sensitivity, Levenshtein similarity had perfect sensitivity and medium specificity. This meant that Levenshtein similarity incorrectly classified 33% of behavioural non-cognates as cognates.

For cognate classification, Levenshtein similarity had an overall accuracy of .742. This was considerably lower than the accuracy of jTRACE. Notably, this accuracy was also lower than the .895 accuracy obtained with the larger set of words in Chapter 3. The optimal threshold obtained for Levenshtein similarity was also higher using the larger set, at 0.310. The restricted subset of words in this chapter may have been insufficient for obtaining a reliable threshold for cognate classification.

The distribution of true positives, true negatives, false positives and false negatives is shown in Table 4.3.

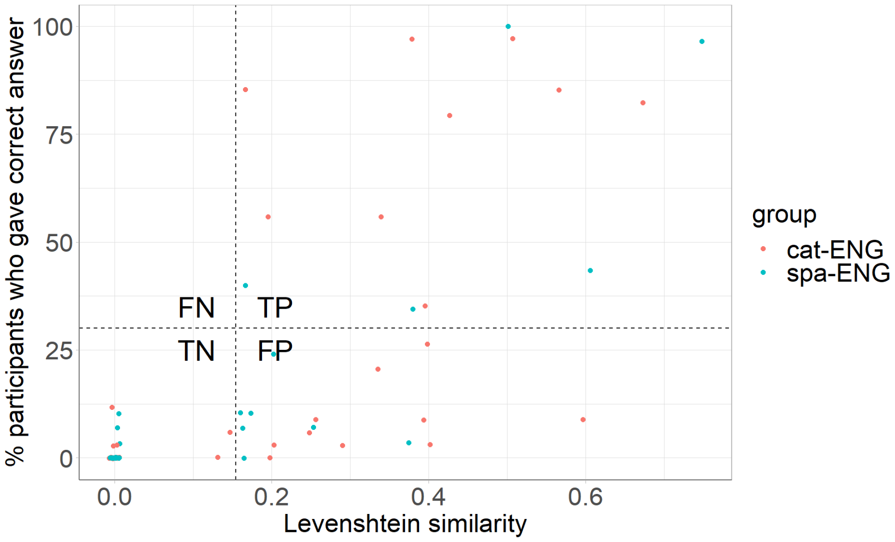

Figure 4.3: Scatterplot of Levenshtein similarity against percentage of participants who gave the correct answer (jTRACE subset), with dotted lines indicating cognate classification cut-offs points and quarters marked for TP, TN, FP, FN

Table 4.3: Distribution of true positives, true negatives, false positives and false negatives (jTRACE subset) when comparing Levenshtein similarity (Lev.) cognate classification against behavioural classification (beh.).

	Lev. cognate	Lev. non-cognate
beh. cognate	14	0
beh. non-cognate	17	35

Performance of ALINE

English-Spanish word pairs had a mean ALINE score of 0.57 (SD = 0.23) and English-Catalan word pairs had a mean of 0.60 (SD = 0.21).

ALINE scores had a correlation of $.646$ with the percentage of participants who provided the correct answer in the translation task (Figure 4.4). The best performance for ALINE according to the Youden method had a threshold of 0.739, with a corresponding specificity of 0.865 and sensitivity of 0.929. The threshold obtained for ALINE using this subset of words was close to the threshold of 0.797 obtained using the larger set in Chapter 3. Translation equivalents pairs with ALINE scores greater than 0.739 were considered cognates, while those with score of 0.739 or less were considered non-cognates. Using this cut-off, there were 20 pairs classified as cognates and 46 pairs as non-cognates. English-Spanish trials had 8 cognates and 18 non-cognates. English-Catalan trials had 12 cognates and 28 non-cognates. Compared to jTRACE, ALINE similarity had lower specificity and higher sensitivity. Importantly, in contrast to the skewed sensitivity and specificity ratio observed for jTRACE and Levenshtein similarity, ALINE had similar values for sensitivity and specificity. This meant that ALINE correctly classified 87% of cognates and 93% of non-cognates. As a result, ALINE had an overall accuracy of .879 for cognate classification. This was lower than the accuracy observed for ALINE in Chapter 3. Importantly, the accuracy of ALINE was lower for the subset of words reported in this chapter than the accuracy of jTRACE.

The distribution of true positives, true negatives, false positives and false negatives for ALINE in comparison to participant responses is shown in Table 4.4.

Table 4.4: Distribution of true positives, true negatives, false positives and false negatives (jTRACE subset) when comparing ALINE cognate classification against behavioural classification (beh.).

	ALINE cognate	ALINE non-cognate
beh. cognate	13	1
beh. non-cognate	7	45

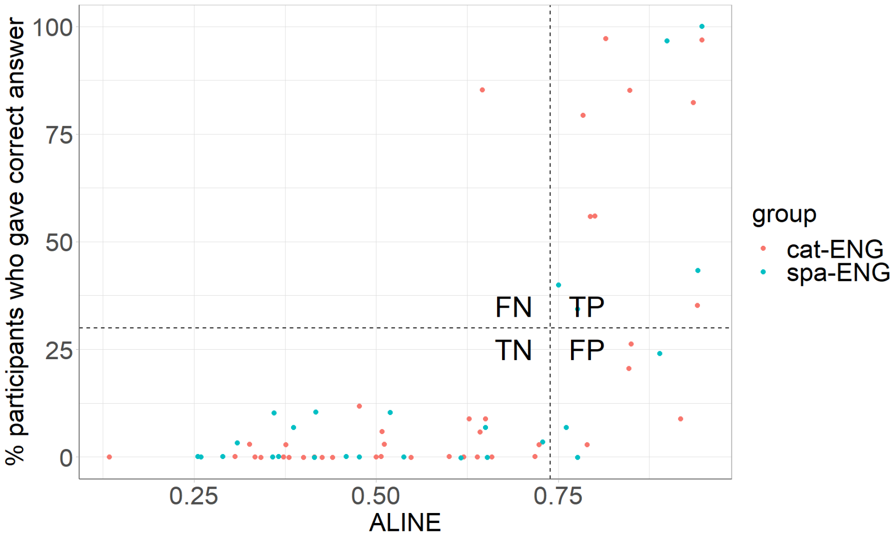

Figure 4.4: Scatterplot of ALINE similarity against percentage of participants who gave correct answer (jTRACE subset), with dotted lines indicating cognate classification cut-offs points and quarters marked for TP, TN, FP, FN

Comparing ROC

The ROC curves for the three measures are visualised in Figure 4.5. In an ROC curve, the sensitivity and specificity corresponding to various possible thresholds is plotted on the same graph, allowing me to gauge the classification performance of a measure even when the threshold is varied. The Area-Under-the-Curve (AUC) is a quantification of the measure's overall performance over different thresholds. The AUC of the jTRACE model was 0.773. The AUC of the Levenshtein model was 0.913. The AUC of the ALINE model was 0.934. Out of these three measures, jTRACE has the smallest AUC, meaning that the predictive success of jTRACE is the lowest among the alternatives. This can be related to jTRACE's low sensitivity. The jTRACE model produced very low scores for some trials that had high participant accuracy, resulting in a high rate of false negatives. Due to the very low activation scores (near to or at the resting value) for this subset of trials, the sensitivity of jTRACE remained low over a wide range of possible thresholds as tested by the ROC analysis.

Interestingly, despite the comparatively low AUC, jTRACE had high accuracy of 0.909 at its optimal cutoff point, exceeding both ALINE and Levenshtein. This can be attributed to its very high specificity of 1, where the model did not incorrectly classify any trials as cognates. This was particularly useful for cognate classification for tested language pairs, as there was a low ratio of cognates relative to non-cognates in the stimuli.

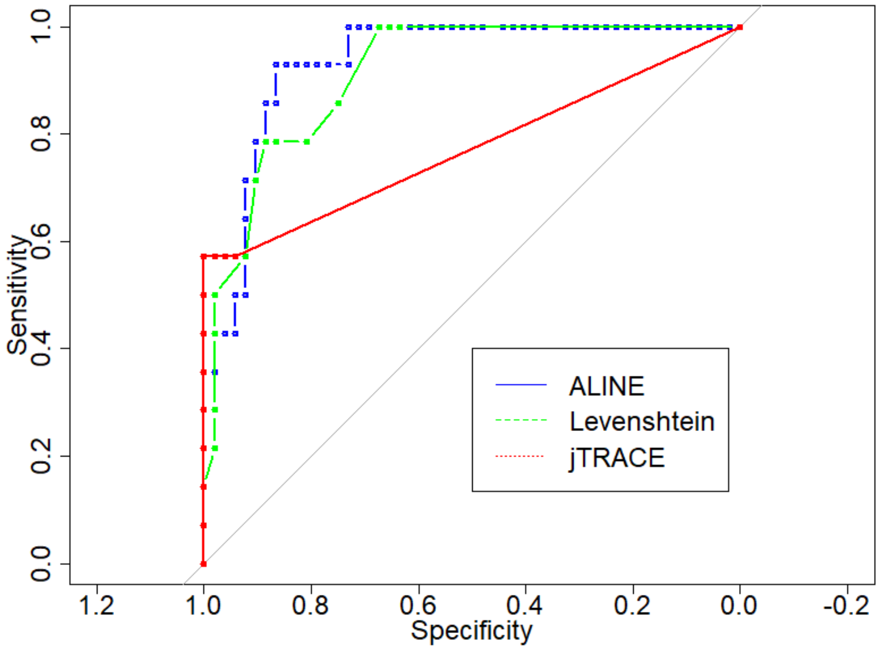

Figure 4.5: ROC curves of jTRACE, ALINE and Levenshtein for predicting cognateness.

All answers

I also investigated whether jTRACE could accurately model the relative probability of participants giving a particular answer to each trial. I took all answers that were given by at least one participant in the translation elicitation task (Figure 4.6). I used the cut-off of 0.234 obtained previously from the ROC analysis. The model had 0.933 accuracy for predicting the percentage of participant answers, with 0.950 specificity and 0.407 sensitivity. Again, the sensitivity of the model was low, with many answers that had high participant response rate receiving very low activation from the model.

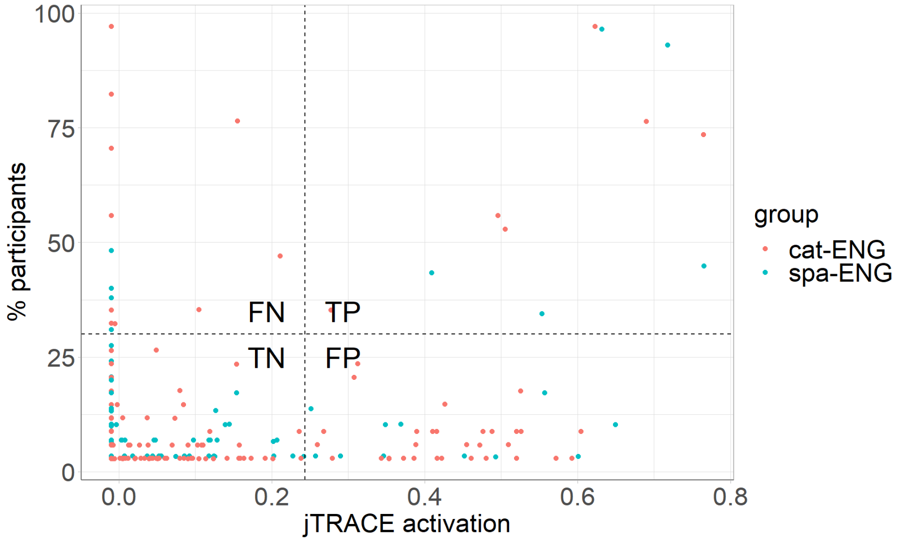

Figure 4.6: Scatterplot of jTRACE activation against percentage of participants who gave a particular answer (jTRACE subset; both correct and incorrect answers), with dotted lines indicating cognate classification cut-offs points and quarters marked for TP, TN, FP, FN. Each point represents a unique answer for a given trial.

I then looked at the answer given by the highest number of participants on each trial, comparing it against the candidate word that received the highest activation in jTRACE. I specifically looked at answers with high consistency (given by at least 30% of participants). Table 4.5 shows the trials where the top participant answer was also the highest activated word in jTRACE.

Table 4.5: List of trials where top participant answer is consistent with the jTRACE candidate word with highest activation.

word	response	% participant	activation
guitarra (gitar@)	guitar (gItAr)	97.1	0.623
guitarra (gitara)	guitar (gItAr)	96.5	0.632
sandwich (sandwiS)	sandwich (sandwItS)	93.1	0.718
pingui (piNgwi)	penguin (pDNgw^n)	76.5	0.690
planta (plant@)	plant (plant)	73.5	0.765
jaqueta (M@kDt@)	jacket (dMak^t)	55.9	0.496
caixa (kaS@)	cash (kaS)	52.9	0.505
pastel (pastEl)	pastel (pastDl)	44.8	0.765
tren (trEn)	train (trEIn)	43.3	0.409
flor (flL)	floor (flLr)	35.3	0.277
ensalada (Ensalada)	salad (sal^d)	34.5	0.553

Trials which were successfully predicted by jTRACE can be characterised by a high proportion of phonological overlap between the input word and candidate word. This supports the utility of jTRACE for identifying phonological similarity.

However, there are other trials where jTRACE produces low activation for the candidate word despite a high proportion of participant selecting the word as their answer. Table 4.6 shows the trials where the top participant answer is inconsistent with the highest activated word in jTRACE.

As with the false negative cognates previously discussed, these trials have moderate phonological overlap as defined using phonological transcriptions. The overlapping phonemes are also non-continuous in many cases. This suggests that par-

Table 4.6: List of trials where top participant answer has low jTRACE activation.

word	response	% participant	activation
xocolata (Sukulat@)	chocolate (tSLkl^t)	97.1	-0.01
girafa (Miraf@)	giraffe (dM@af)	82.4	-0.01
got (gLt)	god (gAd)	76.5	0.155
esponja (@spLnM@)	sponge (sp^ndM)	70.6	-0.01
tigre (tigr@)	tiger (taIg@)	55.9	-0.01
cebra (TEbra)	favourite (fEIv@It)	48.3	-0.01
cuc (kuk)	cook (kUk)	47.1	0.211
chaqueta (tSakEta)	jacket (dMak^t)	40	-0.01
berenar (b@r@na)	banana (b^nan^)	35.3	-0.01
mussol (musLl)	muscle (m^s^l)	35.3	-0.01
tren (trDn)	train (trEIn)	35.3	-0.01
galeta (g@lDt@)	collect (k^lDkt)	32.4	-0.01

ticipants are sensitive to additional cues that are not available to jTRACE's algorithm, for example stress patterns or a consonant bias. I then look at the top jTRACE candidate words for these trials, listed in Table 4.7.

Table 4.7: List of the top jTRACE candidate words for trials where top participant answer has low jTRACE activation.

word	response	% participant	activation
xocolata (Sukulat@)	cool (kul)	0	0.528
girafa (Miraf@)	raffle (raf^l)	0	0.374
got (gLt)	got (gAt)	2.94	0.343
esponja (@spLnM@)	spoon (spun)	0	0.593
tigre (tigr@)	tiara (tiAr^)	0	0.554
cebra (TEbra)	bra (brA)	0	0.359
cuc (kuk)	cock (kAk)	0	0.381
chaqueta (tSakEta)	check (tSDk)	3.33	0.601
berenar (b@r@na)	balloon (b^lun)	0	0.510
mussol (musLl)	moose (mus)	8.82	0.520
tren (trDn)	trend (trDnd)	14.7	0.427
galeta (g@lDt@)	girl (g@l)	2.94	0.572

Candidate words with disproportionately high jTRACE activation can be grouped into two main categories. Some, like “girl” for “galleta” and “bra” for “cebra”, are shorter words embedded within the longer input word. The jTRACE model is capable of separating the input into multi-word strings. For longer foreign words, jTRACE may activate multiple shorter words as top candidates. However, in the translation elicitation task, participants heard words in isolation. While participants were not explicitly asked to give single-word answers, they did so for the majority of trials. The experimental condition of hearing words presented in isolation and being asked to give one answer per trial likely biased them towards assuming the presented words were single words. As such, jTRACE’s automatic tendency to activate shorter words failed to match up to participants’ behaviour for these trials.

In a few cases, the top answer in jTRACE had strong phonological overlap with the presented word, for example “trend” for “tren”, but had low frequency that was the likely reason behind participants’ low rate of responses. This is an issue that could be resolved by adjusting the strength of the frequency parameter in jTRACE. In the default model, there was no affect of lexical frequency applied on word activation. Further exploration is required to obtain the most suitable parameters for modelling frequency effects in participants’ answers.

4.2.3 Study 1 Discussion

All three algorithms had high success for predicting participants’ responses in the translation elicitation task. The accuracy of jTRACE was highest among the three measures. This could be attributed to the perfect specificity of the model's predictions – there were no word pairs that were falsely classified as cognates by the model. There was clear separation between trials considered cognates by the model (activation values above 0.4) and those considered non-cognates (less than 0.1). However, the sensitivity of the model was comparatively low at .571 – almost half of the trials classified as cognates by participant responses were mistakenly classified as non-cognates by jTRACE. The activation values of these false negatives was very low, indistinguishable from trials which were true negatives.

When looking at trials with poor jTRACE predictions, these word pairs tend to have partial non-continuous overlap that results in the candidate word receiving more within-level inhibition than between-level activation, therefore ending the presentation cycles with very low activation activation. This points towards the possibility that the settings chosen for frequency, between-level excitatory effects or within-level inhibition may not be optimal for the task. In this chapter, I used the default settings in jTRACE, which was optimised for monolingual speech recognition. These settings, which are primarily used for modelling the perception of an input word with an exact match in the model's lexicon, may not be the best levels for modelling foreign word speech recognition where there isn't an exact match for the input word in the lexicon. Further work is required to explore the appropriate strength of activation and inhibition in the model for modelling foreign speech recognition.

Another possibility is that the phoneme representations of the model do not accurately represent the degree of similarity between phonemes. If true, this would have spill over effects on the activation from the feature level to candidate phonemes, and also on within-level inhibition between phonemes. I bring up this possibility following the unexpected finding that the English word "owl" (aUl) failed to activate its corresponding word in the lexicon as the top candidate word, instead selecting "eye" (aI) as the top candidate. There was also the issue of jTRACE activating the candidate word "cock" (kAk) for the input word "cuc" (kuk) over the word "cook" (kUk) that is preferred by participants. Some phonemes may be defined as more distant in the model than perceived by listeners and some phonemes may be more similar in the model than perceived by listeners, resulting in inaccurate calculations of word similarity. The second part of this chapter explores the possibility of erroneous phoneme encoding by evaluating jTRACE's accuracy in defining phoneme distance.

4.3 Study 2: Evaluating phonemes activation and inhibition

To function as an accurate measure of phonological similarity, the model must fulfil the following criteria:

1. An input phoneme should provide higher activation to its identical output phoneme compared to any other output phonemes.
2. An input phoneme should provide higher activation to output phonemes that are close in sound (close phonemes) than to output phonemes that do not sound similar (distant phonemes).

4.3.1 Methods

I included all phonemes in Mayor and Plunkett's (2014) expanded set as input phonemes. This comprised of the 10 consonants (/b/, /p/, /d/, /t/, /g/, /k/, /l/, /ɹ/, /s/ and /ʃ/) and 4 vowels (/ʌ/, /a/, /i/ and /u/) from McClelland and Elman's (1986) set and an additional 12 consonants (/m/, /n/, /ŋ/, /v/, /f/, /ð/, /θ/, /ʒ/, /h/, /z/, /j/ and /w/) and 8 vowels (/ɒ/, /ɑ/, /ɛ/, /ɔ/, /ə/, /e/, /ɪ/, /ʊ/) implemented by Mayor and Plunkett (2014). The full list of phonemes with its jTRACE implementations are listed in Appendix E.

Comparing jTRACE phoneme coding against linguist coding

The TRACE model operationalises distance between phonemes using overlap in its feature level. The feature level has an orthogonal representation of each phoneme on 63 units grouped into 7 dimensions (9 units each). The 7 dimensions are Consonantal, Vocalic, Diffuseness, Acuteness, Voicing, Power and Burst. Each unit can have a value of 0 (not activated) or a positive value of 1, 0.5, 0.3 or 0.1 (weighted activation).

I compared the phoneme representation units of jTRACE to a phoneme representation coding system that classified phonemes by a set of distinctive features commonly used by linguists to differentiate phonemes (Karaminis & Scharenborg, 2018). The phoneme representation coding by Karaminis and Scharenborg (2018) had 22 features – Consonant, Vowel, Obstruent, Sonorant, Aspiration, Voice, Plosive, Continuant, Nasal, Lateral, Rhotic, Strident, Labial, Coronal, Dorsal, Glottal, Diphtong, High, Mid, Low, Retracted and Long. For each phoneme, the feature would be coded with 1 if it was a defining characteristic and with 0 if it was not relevant.

I calculated the feature overlap between each non-identical phoneme pair by counting the proportion of units where both phonemes had non-zero values according to jTRACE's feature coding. I then did the same for the linguist coding. This generated a value between 0 and 1 for each phoneme pair for each coding method. I then tested the correlation between the various coding methods.

Comparing jTRACE activation values against ALINE

I also compared the activation values produced by the model for each phoneme. Activation of phonemes is obtained via excitatory connections between the feature level described above and the phoneme level. Output phonemes that have high overlap of activated feature units with the input phoneme receive more activation than output phonemes with low overlap.

As in the word-level evaluation earlier in this chapter, model parameters were left on jTRACE default settings. Phonemes were inserted one at a time to the model in isolation. I extracted the activation levels of all output phonemes at cycle 11 after presentation of each input phoneme, corresponding to the end of the presentation phase for 1 phoneme. I then compared the activation of output phonemes from jTRACE against ALINE similarity scores (Downey et al., 2017) for the same phoneme pair. ALINE scores have a range of 0 (completely different) to 1 (identical).

4.3.2 Results

Phoneme coding

The jTRACE overlap scores had a correlation of .474 (Pearson's *r*) with the phoneme representations by Karaminis and Scharenborg (2018). This correlation is visualised in Figure 4.7. There is medium correlation but also noticeably spread in scores, indicating agreement on some pairs but differing judgements on others.

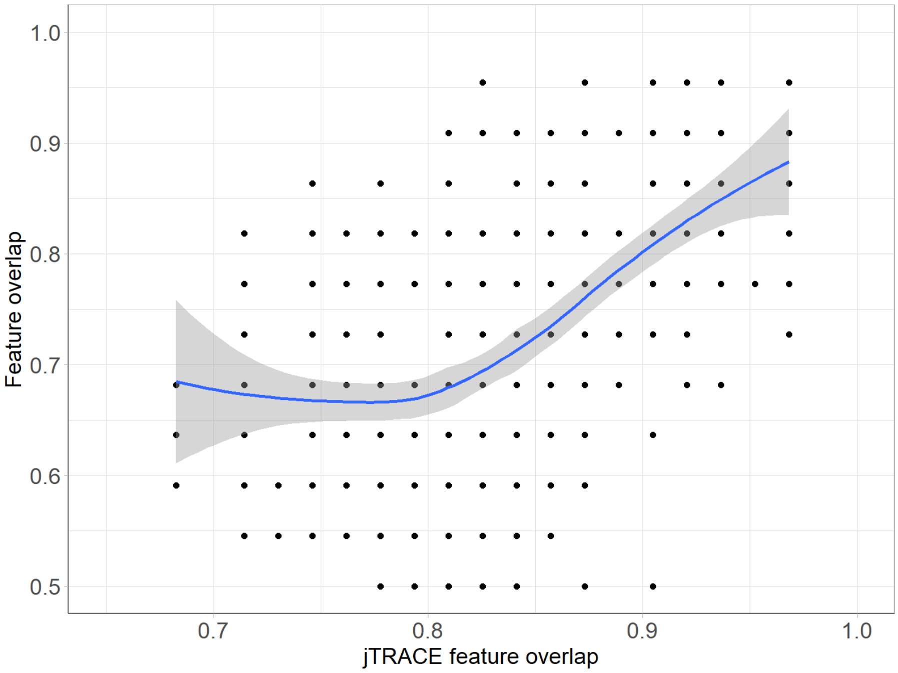

Figure 4.7: Scatterplot of jTRACE feature overlap scores against scores from the coding by Karaminis and Scharenborg (2018) for non-identical phoneme pairs.

jTRACE activation

I then examined the phoneme-level activation in the jTRACE model for both matched phonemes and non-identical phonemes. Activation at cycle 11 for the identical input phonemes ranged from 0.195 to 0.333. The full list of input phonemes (shown with IPA transcription) is shown in Table 4.8 with the activation value at cycle 11 for the identical phoneme, along with the activation of the phoneme that had the second highest activation in response to the input. I see that the identical phoneme has the highest activation in all cases, clearly exceeding the phoneme with second highest activation.

The activation value varied between phonemes even among identical phonemes, with phonemes that had more features activated gaining more activation. Nevertheless, even the lowest activation value of identical phonemes was higher than the highest activation of non-identical phonemes, as seen in Figure 4.8. Activation for non-identical phonemes ranged from -0.100 to 0.191. This fulfilled the first criterion that identical phonemes should always receive the most activation out of all possible output phonemes.

I then explored the second criterion where output phonemes that are phonologically closer to the input phoneme should receive higher activation than output phonemes that are more distant. To test this, I examined the correlation between jTRACE activation and ALINE score. Correlation between jTRACE activation scores for non-identical phonemes had a correlation of $.627$ (Pearson's r) with ALINE similarity score. There was moderate agreement on phoneme similarity between the two measures but there were also many pairs that had noticeably different scores

Table 4.8: Activation level of phonemes in jTRACE as of cycle 11 (corresponding to end of phoneme presentation period), showing activation values of the identical phoneme and the second highest activated phoneme.

input	activation (input)	2nd highest	activation (2nd)
k	0.3330029	g	0.18916203
g	0.3272559	k	0.19105155
t	0.315531	d	0.16765828
p	0.3142206	b	0.16336944
d	0.311724	t	0.16876269
b	0.3099137	p	0.16464504
j	0.2918487	w	0.1288049
l	0.2846597	ɹ	0.17712117
ʃ	0.2766449	s	0.10086848
ɹ	0.2750886	l	0.15978979
h	0.2742904	f	0.10766675
s	0.2726171	ʃ	0.0877165
w	0.2684091	j	0.1005926
ŋ	0.2658726	l	0.09755317
n	0.2648156	m	0.18471743
m	0.250628	n	0.16916232
ʒ	0.2591886	ʃ	0.08947333
v	0.2327064	f	0.14948821
f	0.2314752	θ	0.14812622
θ	0.2292321	f	0.14564742
ʌ	0.2600562	ə	0.16698973
ə	0.240943	ʌ	0.16083148
i	0.2326718	e	0.06316217
ɒ	0.2281001	ɑ	0.14439324
a	0.2163607	ɑ	0.13440206
u	0.2134124	ɑ	0.1214779
ɛ	0.2095283	ɪ	0.1239334
ʊ	0.2013428	e	0.11495326
ɔ	0.2001989	u	0.11402096
ɪ	0.1982358	ɛ	0.11154233
ɑː	0.1962927	a	0.13710134
e	0.1947962	ɪ	0.10773906

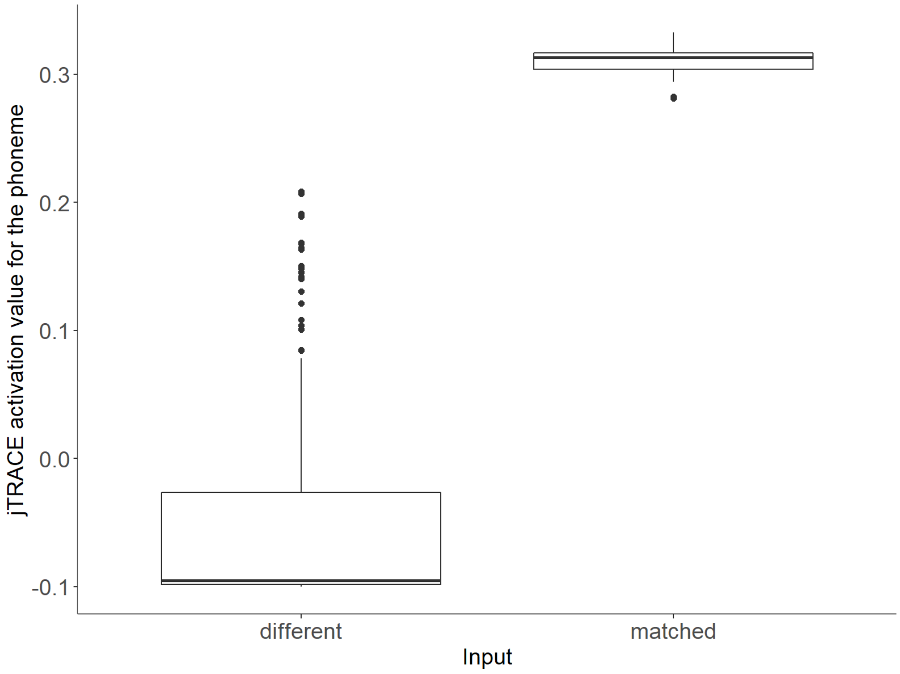

Figure 4.8: Boxplots comparing the mean and range of phoneme activation values for identical and non-identical phonemes.

in the two measures (Figure 4.9).

4.3.3 Study 2 Discussion

At the beginning of this section, I highlighted two main criteria for judging the phoneme representation in jTRACE. I found that jTRACE fulfils the first criteria that output phonemes are activated most strongly by their identical input phoneme. While there is variation in the activation value of identical phonemes, the lowest value is higher than the highest value for non-identical phonemes.

However, there was mixed support for the ratings of phoneme similarity. When

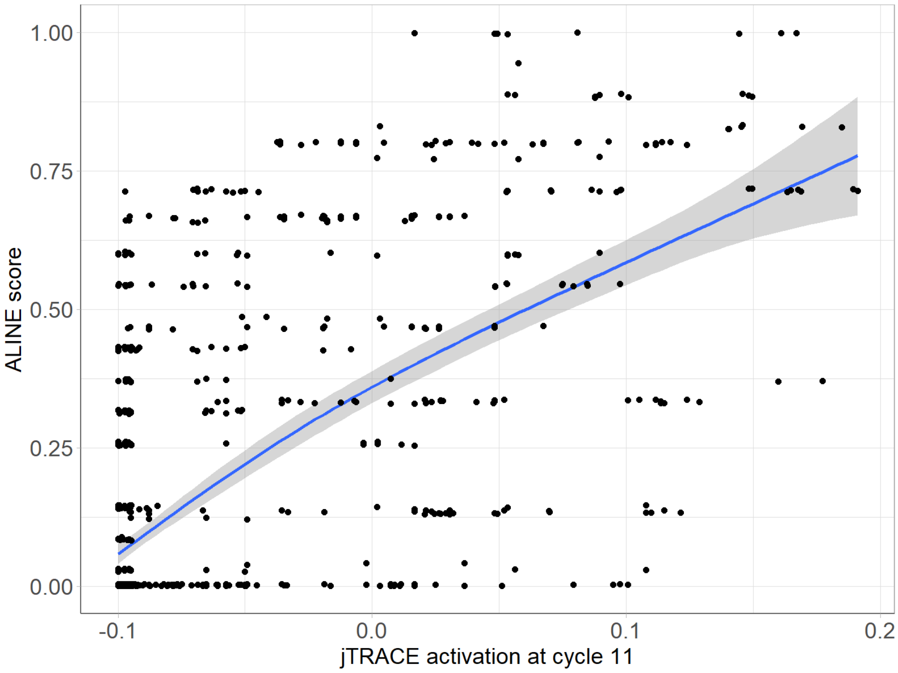

Figure 4.9: Scatterplot of jTRACE activation against ALINE score for output phonemes in response to an input phoneme (non-identical phoneme pairs).

comparing jTRACE against two different methods of calculating phoneme similarity, the correlations were medium in size, suggesting partial but not high agreement between the different methods. While this doesn't irrevocably mean that jTRACE has problems in its phoneme coding (there is the possible that faults lie with the methods it had been compared to), it provides a preliminary exploration into the phoneme representations in jTRACE and the model's ability to model gradient phoneme changes. This analysis would be strengthened with further research comparing jTRACE scores against experimental studies of phoneme similarity. If jTRACE is able to accurately model behavioural judgements of phoneme similarity, it would provide stronger support for the appropriateness of its phoneme representations.

4.4 General Discussion

In this chapter, I evaluated the accuracy of jTRACE for predicting monolingual participants' answers in a translation elicitation. I compared the predictions of jTRACE against similarity scores from Levenshtein distance and ALINE, two algorithms for measuring phonological similarity. While jTRACE had higher accuracy and specificity than the two other algorithms, the sensitivity of the model was particularly poor in comparison. For a subset of trials, there was low activation for the correct answer despite high participant accuracy in the translation elicitation task. In these cases, there were other words that were activated more strongly than the correct answer, subsequently inhibiting the activation of the correct answer back to baseline levels.

I suggested two possible explanations for the unexpectedly low activation for these trials. One possibility could be that the within-level competition effects were weighted too strongly in the model, suppressing possible candidate words. Even if a given candidate has gained activation, mismatches result in it also receiving inhibition, reducing the likelihood of it being the final candidate. Unlike monolingual word recognition, where there is an exact match to the input word in the lexicon, there are no matches in the lexicon for foreign word input. Listeners need to make judgements about the closest equivalent to the input among non-identical options. In many cases, there may be several options with similar overlap available for selection. Accordingly, a model where the word-level inhibition strength is weaker may help the model to maintain more alternatives for selection, instead of quickly suppressing all options aside from the most highly activated one. Another possibility could be that the phoneme representations used in the current study did not provide accurate judgements of phoneme distance. This theory was explored in Study 2 in this chapter. The phoneme distance judgements from jTRACE (as implemented by Mayor & Plunkett, 2014) only had moderate correlation to two other methods of quantifying phoneme distance with which it was compared. It is likely that issues are caused by a combination of the two factors above. If the model judges two phonemes to be distant, it will rapidly start applying inhibition on the mismatched phoneme. Therefore, if the judgements of phoneme distance are inaccurate, the amount of activation and inhibition received by phonemes, and subsequently the words that contain those phonemes, would be likewise inaccurate. A key benefit of the jTRACE interface is that both phoneme representations and strength of within-level inhibition can be flexibly adjusted via its graphical interface. Lexical frequency can also be applied as a modifier for a lexical item's activation value, by affecting a word's resting activation or as a multiplier for phoneme-to-word excitatory connections. If there are multiple candidate words with similar phonological similarity with the input, lexical frequency of the candidate word may play a role in final selection. However, substantial further work is still required to identify the optimal levels for these parameters.

4.4.1 Future directions

Lexicon

For this evaluation, I used a limited lexicon made up of participant answers, to test the model's ability to pick the correct answers among other words given as answers by participants. In the long run, I aim for the model to be usable for identifying cognates without the need to collect behavioural data. Future directions for defining the model lexicon would be to expand it the full list of words from SUBTLEX for studies with adults, or words from the Communicative Development Inventories for studies with infants. The lexical frequency data for the lexicon may also be sourced differently depending on researchers' requirements – SUBTLEX frequencies would be suitable for adult studies and CHILDES frequencies would be more suitable for infant studies. Phoneme sets and model lexicon can be built as data frames in R and converted to jTRACE-readable format using the R package `jtracer` (Garcia-Castro, 2021). It would also be informative to test the model using different language pairs with varying similarity. This would help to validate the model for its accuracy in identifying cognates across language pairs with different proportions of cognates, and also help identifying the optimal cutoff threshold for classifying words as cognate or non-cognates.

Nonnative phoneme perception

One of TRACE's defining aspects that was not explored in this chapter is its interactionist framework, which posits that there is both bottom-up (feature-to-phoneme and phoneme-to-word) activation and top-down (word-to-phoneme) activation during speech recognition. Phonemes can receive different amounts of top-down activation depending on the number of words in the lexicon that contain the given phoneme. Top-down activation effects are interesting to consider in relation to foreign language recognition, as phoneme inventories can both overlap and differ between different languages. It is therefore possible for an input phoneme to not exist in any of the words in the listeners' native language lexicon. In this scenario, while the non-native input phoneme would be able to activate the features associated with it, it would not receive top-down activation from the word level due to a lack of words in the lexicon containing the phoneme. The top-down effects from listeners' native lexicon on spoken word recognition may therefore differ when listening to words in a foreign language as compared to when listening to words in their native language.

Nevertheless, some theories of non-native phoneme recognition posit that a non-native phoneme may be automatically integrated into a existing phoneme category in a listeners' native inventory. Best, McRoberts, and Sithole (1988) put forward a perceptual assimilation hypothesis for non-native phonemes. If two phonemes are assimilated to the same native phoneme category, listeners will find them very difficult to differentiate (single-category assimilation). However, phonemes assimilated to the same category may still be differentiated by how well they fit into the category (category-goodness difference assimilation). If two phonemes are assimilated into two different categories, they will be easily differentiated because they are discriminated on the basis of the assimilated categories (opposing-category assimilation). Finally, if a given phoneme cannot be suitably assimilated into any existing native language phoneme category, it will exist as a new category and thus be differentiable from native phonemes (non-assimilation). The assimilation of non-native phonemes has been supported by experimental findings that listeners consider minimal pairs made from these assimilated non-native phoneme contrasts to be repetitions of the same word (Dufour, Nguyen, & Frauenfelder, 2007; Dobel, Lagemann, & Zwitserlood, 2009). Therefore, the possibility that non-native phonemes may be assimilated into the phoneme inventory of the native language is an important consideration for future expansions of the jTRACE model for foreign word recognition. Parameters for top-down activation in jTRACE may be useful for modelling the effect of lexicon and phoneme set on non-native phoneme recognition.

Salience of overlap

Finally, there may be differences in the salience of phonological overlap. One possibility is that listeners pay more attention to overlapping consonants than overlapping vowels. A consonant bias during word recognition has been identified in adults (Van Ooijen, 1996; Marks, Moates, Bond, & Vasquez, 2002; Creel, Aslin, & Tanenhaus, 2006) and by infants of certain languages (Hochmann et al., 2011; Nishibayashi & Nazzi, 2016). The jTRACE model includes parameters for the activation provided by different feature dimensions. In the default model, all 7 features are weighted equally. Adjusting the weights of the different features may allow us to model the effects of phoneme type, for example by giving stronger weights to features corresponding to consonants.

Another possible source of salience is lexical stress. Phoneme overlap that occurs during a strong or stressed syllable may be weighted more strongly in listeners' judgements than overlap in a weak or unstressed syllable. This may be particularly relevant for languages like English where lexical stress results in differing salience between syllables in a word. Stress patterns may also affect the strength of lexical competition, where competition effects are larger from words that share overlap on the strong syllable. While stress has yet to be implemented in TRACE or jTRACE, the updated version of the Shortlist Model (Norris, 1994) reported in Norris et al. (1995) included parameters to account for lexical stress. Effects of lexical stress may be particularly relevant to the recognition of foreign speech, especially when the native and non-native languages have different stress patterns. Listeners may also be affected by whether the lexical stress matches or is inconsistent across languages (Field, 2005; Van Donselaar, Koster, & Cutler, 2005), though sensitivity to mismatched lexical stress may be affected by listeners' native language (Cooper, Cutler, & Wales, 2002). Future research can consider exploring other models of spoken word recognition to evaluate which model is best for modelling the recognition of foreign words by monolingual listeners.

4.5 Conclusion

This chapter presented an initial exploration into the utility of using a computational model of speech recognition as a measure for cognate definition that combines the key advantages of computed measures and behavioural measures. Using a computational model of speech recognition to identify cognates can help researchers select stimuli with high cross-linguistic similarity and low competition from other similar words in the lexicon. I showed that jTRACE was overall highly successful at modelling monolingual participants' responses in a translation elicitation task, but performed very poorly on a small subset of trials. For successful trials, the model was able to match Spanish and Catalan words to its closest equivalent in the model's English lexicon. However, on some trials, the correct answer had very low activation despite high participant accuracy. I suggested that the low activation may stem from sub-optimal settings for within-level inhibition and between-level activation, along with some issues with phoneme representations. All of these parameters can be adjusted in jTRACE's user-friendly graphical interface, but further work is required to identify the best settings for modelling foreign language recognition. These corrections would need to be done before the model can be extended to include other languages, as was the original intention.

Chapter 5

Cross-linguistic similarity between words facilitates bilingual toddlers' vocabulary trajectories

5.1 Background

Many studies have identified a vocabulary size lag in bilingual children relative to monolinguals, particularly when vocabulary size is calculated using a single language (e.g., Pearson et al., 1993, 1997; Bialystok et al., 2010; Hoff et al., 2012; Cattani et al., 2014). The identification of this lag has led to research looking into strategies

---

This chapter includes elements that have been published with the following reference: Siow, S., Gillen, N. A., Lepadatu, I., Avila-Varela, D., Garcia-Castro, G., Sebastian-Galles, N., & Plunkett, K. (2022). The effect of cognates on bilingual infant vocabulary trajectories: a study using bilingual CDIs of English and one additional language. In Proceedings of the 46th Annual Boston University Conference on Language Development. Cascadilla Press. SS was responsible for data analysis, for helping bilingual children overcome this gap. A unique feature of the bilingual learning environment is that the two languages are interlinked by the existence of words which share meaning. These dual-language word pairs are known as *translation equivalents.* This semantic overlap between languages subsequently gives rise to the question of whether a bilingual's learning process for their two languages are similarly interconnected or largely independent. One way to measure the degree of interdependence between languages is by studying how many translation equivalents are known by the child. This is an approach taken by studies such as Pearson, Fernández, and Oller (1995). The term *doublet* has been used by various studies in the bilingual literature (Pearson & Fernández, 1994; De Houwer et al., 2006) to refer to translation equivalents that a child has learnt in both languages. To give an example, if an English-Spanish bilingual child knows the English word "dog" as well as its Spanish translation equivalent "perro", the child is said to know a doublet. We know that by the second year of life, bilingual children understand words in both of their languages, including many doublets (De Houwer et al., 2006; Legacy et al., 2017). Poulin-Dubois, Bialystok, Blaye, Polonia, and Yott (2013) found that 24-month-old bilinguals who knew more doublets were faster to recognise words in a receptive vocabulary task, even after statistically accounting for the effects of total vocabulary size. Poulin-Dubois et al. linked this finding to the theory by Costa, Miozzo, and Caramazza (1999) that translation equivalents are linked to a shared concept node, and activation of either word sends parallel activation to lexical entries of both translation equivalents, facilitating easier lexical access. For the toddlers in Poulin-Dubois et al. (2013)'s sample, whose lexical representations are still developing, knowing both translation equivalents may also strengthens

---

interpreting results and writing the manuscript.

the activation of the concept node, leading to faster picture recognition in the tested picture-based receptive vocabulary task.

5.1.1 Cognate facilitation effect

Many languages are historically related, and as such may also have words that sound very similar, for example English “train” and Spanish “tren”. TE pairs with shared phonology/orthography and etymology are known as *cognates*. The parasitic model of second language learning proposes that learners of a second language are sensitive to similarities in form to their first language (L1), and are able to make use of these similarities to learn new words in their second language (L2) more easily (Hall, 2002). Bilingual adults have been found to be faster, in both languages, at naming pictures whose names are cognates compared to non-cognates (Costa et al., 2000), as well as being able to retrieve a greater number of cognates in a verbal fluency task (Blumenfeld, Bobb, & Marian, 2016). This suggests that cognates are easier to retrieve from memory than non-cognates. The cognate facilitation effect has been used as support for simultaneous activation of lexical representations from both languages in word recognition and production (Costa et al., 2000). Polish second language learners of English were also found to be better at translating cognates than non-cognates (Otwinowska & Szewczyk, 2017). Additionally, the degree of similarity between cognates affected participants’ translation accuracy, supporting the hypothesis that participants made use of the cross-linguistic similarity to support their task performance. Past research has also found advantages of similarity between languages on vocabulary development in younger bilinguals of different ages. The ‘picture naming’ advantage for cognates found by Costa et al. (2000) was replicated by Poarch and Van Hell (2012) in both adults and school-age (5–8 years old) children. Various studies have found advantages for cognates over non-cognates for children’s performance on receptive vocabulary tasks (Bosma, Blom, Hoekstra, & Versloot, 2019; Pérez et al., 2010) and word production tasks (Bosch & Ramon-Casas, 2014). This cognate advantage has also been found for the acquisition of a second language by adult learners (Otwinowska & Szewczyk, 2017) and school-age children (Tonzar, Lotto, & Job, 2009).

The cognate facilitation effect is posited to operate via cascaded activation between lexical nodes and their phonological segments (Costa et al., 2000). When a cognate word is activated, it also activates its cognate translation equivalent. The lexical nodes feed activation to their corresponding phonological nodes, and the phonemes that overlap across the two words feed activation back to the lexical nodes, heightening the activation of cognate words. This cognate facilitation effect therefore relies on the target word’s translation equivalent having a lexical representation in the learners’ lexicon and being activated. For less proficient language learners and less frequent words, lexical access of L2 words may be mediated by the translation in their L1 (Kroll & Stewart, 1994). They can draw upon similarities between the L2 word and the more familiar L1 word to help process words in the L2 (Kroll et al., 2002). The strength of the translation equivalent’s lexical representation can affect the strength of the cognate facilitation effect. The variable effects of cognates on L1 and L2 word recognition and learning will be discussed in the next two sections.

5.1.2 L2 proficiency and cognate facilitation

Much of the literature on the cognate facilitation effect has presented evidence of cognate facilitation on word recognition and word learning in an individual's second language. This facilitatory effect of cognates is typically more prominent when individuals are tested in their less proficient language as compared to their first language. Adult participants have been found to have a strong cognate facilitation effect when asked to name pictures in their L2 (Costa et al., 2000; Poarch & Van Hell, 2012). Kroll et al. (2002) found a benefit of cognateness on translation speed where cognates were translated faster than non-cognates, both when translating from L1 to L2 and from L2 to L1. This was true for both fluent and less fluent L2 learners, however the cognate advantage was particularly strong for less fluent learners. Several other studies have also found stronger cognate facilitation effects in low-proficiency L2 learners compared to high-proficiency L2 learners when tested in the L2, in lexical decision (Van Hell & Dijkstra, 2002), reading performance (Bultena et al., 2014) and performance on a receptive vocabulary test (Allen, 2019). This effect of cognateness on L2 has also been found in children. In 3–5-year-old English-Spanish bilinguals, the size of the cognate facilitation effect in English was positively correlated to bilinguals' Spanish dominance, with children who hear more Spanish showing stronger cognate facilitation when tested with an English receptive vocabulary task (Robinson Anthony, Blumenfeld, Potapova, & Pruitt-Lord, 2020). In a study with 5–7 year old English-Spanish bilinguals, Pérez et al. (2010) found that in contrast to English-dominant bilinguals showing better performance for non-cognates than cognates in an English receptive vocabulary task, Spanish-dominant bilinguals showed better performance for cognates than non-cognates, supporting the stronger manifestation of cognate facilitation in the L2. Kindergarten and school-age children learning two dialects of Arabic were found to have stronger phonological representations for cognates than non-cognates when tested in the later-learned dialect (Saiegh-Haddad & Haj, 2018). Identical cognates exhibited the strongest effect, and there was an advantage for partially-overlapping cognates over non-cognates. Taken together, these studies suggest that phonological overlap between languages can facilitate the acquisition of L2 words and the strength of its encoded phonological representation. Bosma et al. (2019) likewise found that bilingual children with low exposure to Frisian benefited the most from cognates when tested on receptive vocabulary performance in their non-dominant language. This suggests that children can compensate for poorer knowledge in their non-dominant language by making use of knowledge of their dominant language. Explicit awareness of cognates between L1 and L2 has been positively associated to word reading, vocabulary and reading comprehension in the L2 in students immersed in the second language when tested at Grade 1 and 2 (Hipfner-Boucher, Pasquarella, Chen, & Deacon, 2016). The cognate awareness measure tested students' ability to identify cognates and reject both false cognates and non-cognates. In summary, the discussed findings of cognate facilitation in L2 test performance for both adults and children suggests that learners can make use of cross-linguistic similarity between their languages to support word learning and retrieval in their weaker language. Learners subsequently become less dependent on this cross-linguistic similarity as their second language proficiency grows.

However, seemingly in contradiction to the research showing that more proficient L2 speakers show less reliance on cognates on L2 language performance, several of the aforementioned studies with children also found an *increase* of the cognate facilitation with age. In Bosma et al. (2019)'s study, the observed cognate facilitation effect increases with age, suggesting that children become more sensitive to form similarity between languages as they get older and gain more language experience. In another study with 8–15-year-old children, reliance on cognates for reading comprehension in the L2 is also stronger in younger children (who are less proficient in reading) than older children (who have grown more familiar with the printed word) (Duñabeitia, Ivaz, & Casaponsa, 2016). So how do we reconcile this pattern of increased cognate facilitation with age but decreased cognate facilitation with higher proficiency? Van Hell and Tanner (2012) suggested that cognate facilitation effects in the test language may be only manifest if learners have achieved a minimum level of proficiency in the non-test language. This theory will be discussed in the following section.

5.1.3 L1 proficiency and cognate facilitation

Cognate facilitation in translation and picture naming have been found even in highly-proficient bilinguals and trained interpreters (Christoffels, De Groot, & Kroll, 2006), supporting theories of the simultaneous activation of both languages even in task conditions that only target one language. In Costa et al. (2000) and Poarch and Van Hell (2012)'s studies, cognate facilitation was found in L1 picture naming as well, though the strength of the effect was smaller compared to L2 picture naming. However, it is important to note that cognate facilitation in the L1 was only found in high-proficiency bilinguals and not low-proficiency L2 learners (Poarch & Van Hell, 2012). Poarch and Van Hell found that German children (5–8 year old) learning English in school as an L2 (average 1.50 years English immersion) only showed cognate facilitation when tested in picture naming in their L2 (English) but not when tested in their L1 (German), while balanced child bilinguals (exposed to both languages at home, average 2.67 years immersion) showed cognate facilitation in both English and German. Similarly, in another study with school-age Dutch-English bilingual children (10, 12 and 14 years old), all three age groups showed faster recognition for cognates than non-cognates in a lexical decision task when tested in their L2 (English) but not when the lexical decision task was administered in their L1 (Dutch) (Brenders, Van Hell, & Dijkstra, 2011).

These findings support the idea that the cognate facilitation effect for performance in the test language only manifests to a noticeable extent when participants have sufficient proficiency in the non-test language. This is consistent with the fuzzy phono-lexical representations hypothesis put forward by Cook, Pandža, Lancaster, and Gor (2016) for learners of a second language. Lower-proficiency L2 speakers are worse at differentiating phonologically-similar words even when the words do not contain difficult phonological contrasts, leading Cook et al. to posit that non-native phono-lexical representations are less detailed and thus more prone to be confused with other similar words. Cook et al. also found that low proficiency L2 learners are more vulnerable to interference from phonologically-related competitors when attempting to retrieve the lexical representation of an L2 word as compared to high proficiency speakers. This fuzzy phono-lexical representation and subsequent difficulty in retrieving the associated semantic content may explain why cognate facilitation from L2 to L1 is not found in low proficiency L2 learners for tasks involving word recognition and production.

What does this mean for simultaneous bilingual toddlers, whose lexical representations are still developing in early learning? If the cognate facilitation effect is dependent on an existing lexical representation having a minimum strength, would toddlers be able to make use of cognates in their word learning? There is evidence that even toddlers are sensitive to cross-linguistic similarity across their languages. Even 18–53-month-old English-German bilingual toddlers showed the asymmetry of the cognate facilitation effect found in older children and adults, with strong phonological overlap facilitating word recognition in bilinguals' L2 (English) but no significant effect on word recognition in L1 (German) (Von Holzen et al., 2019). Studies of vocabulary size trajectories in bilingual toddlers have also provided support for an effect of cross-linguistic similarity on young bilinguals' word learning. Floccia et al. (2018) measured the vocabulary of 24 month old bilingual toddlers using vocabulary questionnaires, finding that children learning typologically closer languages had larger vocabularies than those learning more distant ones. In a similar study, Gampe, Quick, and Daum (2021) studied toddlers (18 to 30 months) learning Swiss German and one other language, finding larger vocabularies in those whose other language was more similar to Swiss German.

5.1.4 The Present Study

I am interested in the cognate facilitation effect in early vocabulary acquisition of simultaneous bilingual toddlers. Studying this effect can tell us about young bilingual learners' ability to compare their two languages and find similarities between them. Evidence of cognate facilitation in toddlers' vocabulary trajectories would lend support to theories of simultaneous bilingual lexical activation even at this early stage of language development. I explore the hypothesis that cognate facilitation occurs when learners compare a novel word form to an existing lexical representation. If the effect requires comparison to existing representations, I would expect the cognate effect to be stronger in a toddler's less-dominant language, and found only when the toddler is also reported to know the word in the other language.

To study this, I collected data online from bilingual families with children between 12- and 36-months old. These families spoke English and one other language. Parents filled in two vocabulary questionnaires, one in English and one in the other language. The vocabulary questionnaires used in the present study were deliberately formulated to have high conceptual overlap, allowing me to directly compare reported vocabulary knowledge in toddlers' two languages. By analysing only the subset of concepts that overlap across all our vocabulary questionnaires, I am able to study the effect of cross-linguistic word similarity on the same list of concepts, thus avoiding the possible confound that observed differences in vocabulary knowledge may be caused by variation in concept complexity across groups. Some families contributed longitudinal data, answering the same vocabulary questionnaire across multiple timepoints.

Analyses on the impact of cognates on word knowledge will be conducted on cross-sectional data in Study 1, and supported by longitudinal data in Study 2. Finally, Study 3 will include analyses on how the percentage of cognates across languages may affect toddlers' vocabulary size for toddlers learning different language pairs.

5.2 Study 1: Cross-sectional

The aim of Study 1 is to investigate whether cognates facilitate vocabulary learning in this population of bilingual toddlers, and whether toddlers' language dominance and existing lexicon affects the strength of the cognate facilitation effect. The research questions are as follows:

1. Are words that are cognates more likely to be known than non-cognates after accounting for word difficulty, child's age and child's language exposure?
2. Is the cognate facilitation effect stronger for bilinguals in their less dominant language?
3. Is the cognate facilitation effect dependent on knowledge of the translation equivalent (TE)?

5.2.1 Methods

Participants

The full sample consisted of 12 to 36-month-old bilingual toddlers (N = 778; N female = 398, N male = 379, N other = 1) (age 12.0–36.0, mean 24.0 months) growing up with English and one other language (Dutch, French, German, Italian, Polish, Portuguese or Spanish). This sample includes the UK sample reported in Chapter 2 (N = 622), with additional families with primary residence in Germany (N = 48), the Netherlands (N = 28) or Spain (N = 80) at the time of the study. Similar to the UK sample, recruitment was conducted online via social media (Facebook, Instagram) for the samples from Germany, the Netherlands and Spain. Participants who completed the questionnaires were offered a €5 Amazon voucher as thanks for completing the study.

As all of the bilinguals had the common language of English, I will henceforth follow the convention used by Floccia et al. (2018) and refer to the non-English language as the additional language (AL). Some participants also contributed longitudinal data—this will be analysed in Study 2. Participants received a £5 Amazon voucher or t-shirt (UK sample) or a €5 Amazon voucher (Germany/Netherlands/-Spain sample) for each time they completed a longitudinal questionnaire that was sent to them. For participants who contributed longitudinal data, only the first data point was used for the cross-sectional analyses in Study 1, while the longitudinal analyses in Study 2 only includes the subsequent data points.

I collected demographics information about each child's age, gender, parent education level, premature birth and hearing problems. Participants were excluded if they were born more than 6 weeks premature and/or had hearing problems. Overall, socioeconomic status as judged by parent education level was high, with 89.0% of mothers reported to hold a University degree or equivalent. To obtain an estimate of the bilingual language environment, I used a language exposure questionnaire adapted from the questionnaire used by Bosch and Sebastián-

Gallés (2001). Items included in this questionnaire can be found in Appendix A. Of their two languages, bilinguals in the sample were exposed at least 20% of the time to each language (and therefore no more than 80%). Children who heard 10% or more of a third languages were excluded as trilinguals. All children in the UK sample have at least 1 parent who is a native speaker of the AL, while all children in the Germany/Netherlands/Spain samples have at least 1 parent who is a native speaker of English. In this manner, all children are exposed to native input of the non-community language in the home.

Vocabulary questionnaires

I collected parent-report vocabulary questionnaires in English and the AL for each child. Parents mark whether their child understands and says, understands but does not say, or does not understand each word in the list. For English, I used the Oxford Communicative Development Inventory (CDI) (Hamilton et al., 2000). The list of words and categories can be found in Appendix B. The Oxford CDI is a vocabulary questionnaire for British English infants, normed using data from monolingual infants (12–25 months old). The Oxford CDI was chosen for use due to the UK demographic of the present study. Other bilingual studies conducted in the UK have similarly used the Oxford CDI (or its short form) to measure English vocabulary knowledge (Cattani et al., 2014; Floccia et al., 2018).

Vocabulary questionnaires have been widely used in bilingual research (Pearson et al., 1993; Pearson & Fernández, 1994; Cattani et al., 2014; Floccia et al., 2018; O'Toole et al., 2017; O'Toole & Hickey, 2017). Such vocabulary checklists have been used to compare the vocabulary growth of bilinguals both for children learning the same language pair and across children learning different language pairs. Various methods have been used when comparing across different language pairs, as the number of words often differs between adaptations. O'Toole et al. (2017) calculated z-scores for conceptual vocabulary sizes as a proportion of the total possible concepts for each language pair. Using a different approach, Floccia et al. (2018) compared the subset of concepts that overlap across all the languages in their sample. Both approaches have their pros and cons. While O'Toole et al. (2017)'s method allowed a larger subset of words to be analysed, the scores are sensitive to the average word difficulty of different questionnaires. On the other hand, while the approach taken by Floccia et al. (2018) ensures that the words are directly comparable in the concepts sampled, the degree of overlap between some adaptations may be relatively small (Floccia et al. had only 30 words that overlapped across all their questionnaires), limiting the statistical robustness of the analyses.

For the present study, I chose to use translations of the Oxford CDI to measure vocabulary knowledge in each AL, in order to to maximise the conceptual overlap between questionnaires. For each AL, a CDI was adapted by translating the words in the Oxford CDI. Native speakers of each language were consulted for the creation of these adaptations. Not all words in these adaptations were direct translations of the Oxford CDI words – this is due to some cultural differences and linguistic differences requiring us to substitute some words with more appropriate replacements. For a small number of words, an item in the Oxford CDI was deemed not culturally relevant or not directly translatable to the AL. Meanwhile, linguistic constraints were cases where two Oxford CDI entries of the same category had the same word as their translation. For example, both “clock” and “watch” are items in the Oxford CDI under the “household items” category, but both words are translated as “Uhr” in German. Rather than keeping both in the German adaptation of the CDI, and thus potentially confusing respondents, I chose to replace one of them – I kept the more frequent word, and replaced the less frequent one with another one in the same semantic category. In cases where the words were of a different category, as in the animal “fish” and food item “fish”, I kept both as the referent for each entry was clear. For the analyses reported in this chapter, I used the subset of concepts where translations were present in all our CDIs. The list of common concepts can be found in Appendix C with the translations in each language listed exactly as they were presented as entries in the respective CDIs.

Sample subset for vocabulary in comprehension

It is important to considering the constraints of the methodology when conducting research using specific tool. The Oxford CDI is comprised of a finite list of words aimed to collect and compare infants’ vocabulary knowledge over the second and third year of life. It is not meant to be a comprehensive list of all words children could possibly know. As a result, the CDI is naturally prone towards floor and ceiling effects, where the youngest infants may know none of the words in the CDI, and older toddlers may know all the words (and likely more words as well). For the research questions asked in this chapter, which investigate the developmental trajectories of different words, data from toddlers who know none or all of the words in the CDI does not provide me with the necessary variation needed to test my hypotheses. Therefore, I needed to avoid these floor and ceiling effects

Table 5.1: Number of participants per language (comprehension sample), with mean age at time of response and mean percentage of overall English exposure.

language	country	N	mean age	mean Eng exp (%)
Dutch	Netherlands	21	28.3 (4.91)	52.6 (15.6)
Dutch	UK	27	27.7 (4.70)	63.0 (15.8)
French	UK	63	27.5 (4.91)	62.1 (17.5)
German	Germany	37	28.4 (4.90)	40.1 (17.1)
German	UK	37	28.3 (3.92)	59.7 (16.3)
Italian	UK	68	26.7 (4.48)	57.2 (16.0)
Polish	UK	62	28.0 (4.57)	53.3 (18.8)
Portuguese	UK	42	27.7 (3.75)	48.9 (19.2)
Spanish	Spain	53	28.8 (4.26)	51.3 (15.8)
Spanish	UK	121	27.7 (4.85)	53.5 (16.3)

on the analyses. Analyses on vocabulary knowledge in comprehension used a subset of participants from the original 12 to 36-month-old sample of bilingual toddlers. I employed the same method of deciding cut-offs as in Chapter 2. I defined the ceiling as 90% of the maximum conceptual vocabulary size as calculated from overlapping concepts in the questionnaires, and the floor as 10% of the maximum conceptual vocabulary size. I excluded age groups (rounded by age in months) that had median conceptual vocabulary sizes in comprehension smaller than the floor value and those that had median conceptual vocabulary sizes larger than the ceiling value. The sample subset for analyses for vocabulary size in comprehension was 12 to 26-months-old. Only bilinguals were analysed in this chapter, so the monolingual sample's cut-off points were not considered here, making the age range slightly different from that reported in Chapter 2. The number of toddlers in the comprehension sample was 464 (N female = 236, N male = 228). Distribution of the sample for comprehension (cross-sectional), split by toddlers' AL and country of primary residence is shown in Table 5.1.

Table 5.2: Number of participants per language (production sample), with mean age at time of response and mean percentage of overall English exposure.

language	country	N	mean age	mean Eng exp (%)
Dutch	Netherlands	14	22.5 (3.23)	51.7 (16.6)
Dutch	UK	19	23.5 (3.48)	59.5 (16.5)
French	UK	61	22.5 (3.30)	60.7 (19.1)
German	Germany	34	23.3 (3.25)	38.2 (15.0)
German	UK	30	24.3 (3.40)	60.7 (16.6)
Italian	UK	72	23.1 (3.41)	53.2 (15.6)
Polish	UK	74	22.7 (3.34)	52.5 (18.6)
Portuguese	UK	29	23.3 (3.64)	54.8 (18.4)
Spanish	Spain	55	23.4 (3.51)	44.3 (17.5)
Spanish	UK	139	22.7 (3.16)	50.1 (17.7)

Sample subset for vocabulary in production

I similarly aimed to avoid floor and ceiling effects for vocabulary knowledge in production. As with comprehension, the ceiling was defined as 90% of the maximum conceptual vocabulary size, and the floor as 10% of the maximum conceptual vocabulary size. I excluded age groups that had median conceptual vocabulary sizes in production smaller than the floor value and those that had conceptual English vocabulary sizes larger than the ceiling value. The sample subset for analyses for vocabulary size in production was 20 to 36-months-old. The number of toddlers in the production sample was 531 (N female = 275, N male = 255, N other = 1). Distribution of the sample for production, split by toddlers' AL and country of primary residence is shown in Table 5.2.

Defining word difficulty

To obtain a proxy for English word difficulty for word-level analyses, I used monolingual age-of-acquisition (AoA) norms obtained from 1720 British English monolingual infants between 12- and 32-months old. Vocabulary knowledge data was collected using the Oxford CDI (Hamilton et al., 2000), inclusive of data that is publically available on Wordbank (Frank, Braginsky, Yurovsky, & Marchman, 2017). A toddler's age is rounded to the nearest month. AoA for a word is defined as the earliest age (in months) that the word is known by at least 50% of toddlers of that age in this sample. This measure provides the predicted developmental trajectory of words in the absence of bilingual language interaction.

Defining cognateness

In Chapter 3, I showed that ALINE similarity performed slightly better than Levenshtein similarity for classifying cognates. ALINE similarity was therefore used in this chapter. To define phonological similarity between words, I calculated the ALINE score (Kondrak, 2003) between the phonological transcriptions of 306 word pairs common across all CDIs, using the R package `alineR` (Downey et al., 2017). ALINE calculates the graded phoneme distance between phonemes. The similarity between two phonemes is calculated according to the degree of similarity between the phonemes' defining features (e.g., manner of articulation, voicing, etc; see Kondrak, 2003 for details). The `alineR` package provides a standardised score which divides the total score by the length of the words, producing a score between 0 (maximally dissimilar) and 1 (identical). I defined cognates as word pairs with a standardised score of more than 0.797, and non-cognates as pairs with score of 0.797 or more (following the optimal cut-off identified in Chapter 3. For the analyses in this chapter, identical cognates (i.e. cognates with the same pronunciation across languages) were excluded, as identical cognates cannot be reliably be assigned to a specific language because the word forms are identical across languages. Onomatopoeia were also excluded. After exclusions, 306 common concepts across all languages were included in the analyses reported in this chapter.

Dutch (20.3% cognates) and German (16.0%) shared the largest number of cognates with English by a large margin. This was followed in order by French (7.8%), Italian (7.8%), Spanish (7.8%), Portuguese (5.6%) and Polish (4.3%).

Analysis Plan

To test the hypothesis that cognates will be learnt earlier than non-cognates, I ran 2 binomial generalised linear mixed effects models with the following binary categorical dependent variables derived from the vocabulary questionnaires respectively:

1. English comprehension (understands or does not understand in English)
2. English production (produces or does not produce in English)

I included five main predictors in each model:

1. the word pair's cognateness (binary variable)

2. the word's difficulty (operationalised using AoA in months derived from English monolingual data, with words that have later AoA defined as being more difficult, range 12–32 months, centered on mean and scaled by SD)
3. AL TE knowledge (whether or not the toddler knew the translation equivalent in their AL)
4. the toddler's age in months (centred on mean age and scaled by SD)
5. the bilingual toddler's percentage of English exposure (centred on 50% and scaled by SD)

I also included 2 interactions in each model. To test the hypothesis that cognate facilitation only manifests when there is existing knowledge in the other language, I included an interaction between AL TE knowledge and cognateness. To study if the cognate facilitation is stronger in the less dominant language, I included an interaction between cognateness and toddler's English exposure. If cognate facilitation benefits the weaker language more than the stronger language, I would expect to see a negative interaction between these variables for English vocabulary (i.e. toddlers with less English exposure showing stronger cognate facilitation in English) and a positive interaction for AL vocabulary (i.e. toddlers with more English exposure showing stronger cognate facilitation in AL).

Participant and word were included as random effects.

The model syntax for the full model is shown below:

```
glmer( ~ age + English_exposure + word_difficulty
    + AL_word_knowledge + cognateness +
```

```
AL_word_knowledge:cognateness +
English_exposure:cognateness + (1|participant)
+ (1|concept))
```

Model comparisons were conducted for the following models of differing complexity in fixed effects:

1. Model 0: age + English exposure + word difficulty
2. Model 1: age + English exposure + word difficulty + AL TE knowledge
3. Model 2: age + English exposure + word difficulty + AL TE knowledge + cognateness
4. Model 3: age + English exposure + word difficulty + AL TE knowledge + cognateness + AL TE knowledge:cognateness
5. Model 4: age + English exposure + word difficulty + AL TE knowledge + cognateness + AL TE knowledge:cognateness + English exposure:cognateness

Each model was compared to the one directly preceding it to test the unique variance explained by each new predictor added. The random effects of all models was kept constant. These comparisons were conducted separately for English comprehension (Models C0, C1, C2, C3, C4) and English production (Models P0, P1, P2, P3, P4). To evaluate the models, I used marginal R^2 and the Akaike Information Criterion (AIC). Marginal R^2 (R^2m) was obtained using the `MuMiN` package in R, which uses the delta method for deriving observation-level variance of a model relative to a null model with only the intercept term. The AIC of each model, Chi-square statistics comparing models and p-values for Chi-square were obtained using the `anova` function in base R.

5.2.2 Results

English comprehension

Figure 5.1 shows the predicted likelihood for the binomial generalised linear model with English comprehension as the dependent variable and toddler's English exposure, word difficulty and word pair cognateness as predictors, including an interaction between English exposure and cognateness. Firstly, we see the expected trend that English words that are more difficult are known by less bilingual toddlers. English words that are cognates with their AL translation equivalent are known by more toddlers compared to non-cognates, but only when language dominance is controlled for. This is reflected in the model coefficients listed in Table 5.3, where the main effect of cognateness is actually negative. This means that before taking into consideration language dominance and prior word knowledge, the sampled cognates were more difficult to learn than non-cognates. Additionally, toddlers who have lower percentages of English exposure show a greater difference between cognates and non-cognates than bilinguals with higher percentages of English exposure.

Figure 5.2 shows the predicted likelihood for the binomial generalised linear model with English comprehension as the dependent variable and toddler's knowl-

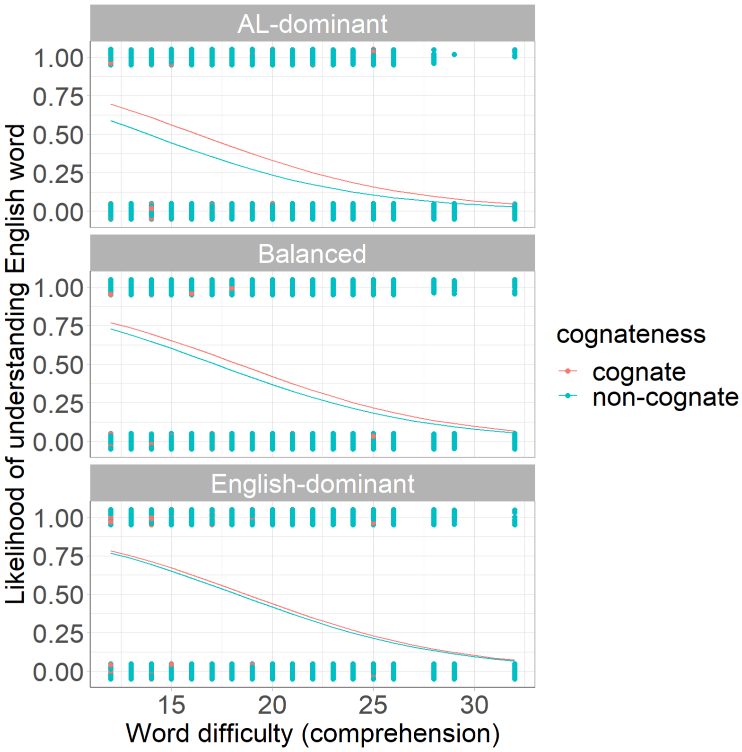

Figure 5.1: Plot visualising model predictions for the generalised linear model of English comprehension, with lines of best fit showing the probability of each word being understood depending on its word difficulty. Points represent knowledge of individual words for each child, where 1 indicates that the word was understood and 0 indicates that the word was not understood. The plots are split by cognateness and faceted by English exposure.

edge of the AL translation equivalent, word difficulty and word pair cognateness as predictors, including an interaction between AL TE knowledge and cognateness. The cognate facilitation effect is larger for words where the toddler also understands the TE.

Coefficients for the English comprehension full model (Model C4) are shown in Table 5.3. P-values were obtained using lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017) via Satterthwaite's approximation degrees of freedom. We see a pos-

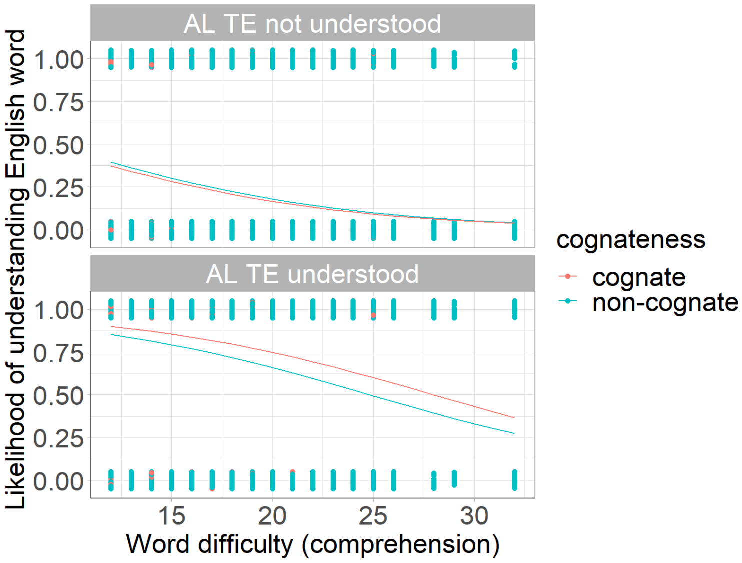

Figure 5.2: Plot visualising model predictions for the generalised linear model of English comprehension, with lines of best fit showing the probability of each word being understood depending on its word difficulty. The plots are split by cognateness and faceted by whether or not the child understands the AL TE.

itive main effect of age, indicating that older toddlers are more likely to know a given English word with increasing age. The positive main effect of English exposure indicates that toddlers who hear more English are more likely to know a given English word than those who hear less English. The negative main effect of word difficulty indicates that words that are more difficult are less likely to be known. The positive main effect of AL TE comprehensions indicates that it is more likely for a toddler to understand an English word if they also understand the TE in their AL. The negative main effect of cognateness indicates that cognates are less likely to be known than non-cognates. While the main effect is seemingly in contrast to my predictions, the positive interaction between AL TE comprehension reveals an interesting pattern—cognates are more likely to known than non-cognateness only when the toddler also understands the TE in their AL. This supports my hypothesis that the cognate facilitation effect is dependent on AL word knowledge. Finally, the negative interaction effect between cognateness and toddler's English exposure indicates that toddlers who hear less English show a larger difference between cognates and non-cognates for English comprehension.

Table 5.3: Model coefficients for the full generalised linear mixed effect model of English comprehension (Model C4), with age in months, English exposure, AL TE knowledge, word difficulty and cognateness as predictors, an interaction between English exposure and cognateness, and an interaction between AL TE knowledge, and cognateness.

Predictor	Estimate	Std Error	z	p
(Intercept)	-1.554	0.082	-18.92	$<$.001
Age	1.232	0.076	16.112	$<$.001
Word difficulty	-1.059	0.029	-35.92	$<$.001
English exposure	0.603	0.077	7.837	$<$.001
AL TE comprehension	1.693	0.020	83.067	$<$.001
Cognateness	-0.198	0.049	-4.014	$<$.001
Cognateness : AL TE comp.	0.668	0.061	10.892	$<$.001
Cognateness : English exposure	-0.063	0.031	-2.038	.0415

Comparisons of models for English comprehension with increasing complexity in fixed effects are shown in Table 5.4. The main predictors of cognateness and AL TE knowledge, the interaction between cognateness and AL TE knowledge, and the interaction between cognateness and English exposure all added significant variance to the model when added one at a time. However, the interaction of cognateness and English exposure has a conflicting result – while the addition of the interaction improved the AIC, the marginal R^2 actually decreased slightly relative to the simpler model. Given that the AIC improvement was very small and there was a decrease in marginal R^2, I interpret this as meaning that the interaction with English exposure did not explain unique variance in a model that already includes the interaction between cognateness and AL TE knowledge.

Table 5.4: Model comparisons for English comprehension. Change statistics compare each model with the one directly above it.

Model	R^2m delta	R^2m change	AIC	Chisq	Pr(>Chisq)
C0	.357		107605		
C1	.424	.0674	99160	8447.6	$<$.001
C2	.424	.000221	99153	8.45	.00365
C3	.425	.00103	99019	136.6	$<$.001
C4	.425	$-$.0000377	99017	4.12	.0427

English production

Similar patterns were found for English production. Figure 5.3 visualises the effect of English exposure on the cognate facilitation effect. As with comprehension, the cognate facilitation effect is stronger in the less-dominant language for English production. Figure 5.4 visualises the effect of AL TE production on the cognate facilitation effect. Again, the cognate facilitation effect only manifests for words where the AL TE is also produced by the same toddler. This is supported in the model estimates, where the main effect of cognateness was not significant, supporting the idea that cognates are not inherently easier to learn than non-cognates before knowledge of the TE in the other language is taken into account (Table 5.6).

The model comparisons are shown in Table 5.5. As with comprehension, the main predictors of cognateness and AL TE knowledge, along with the interaction between cognateness and AL TE production added significant variance to the model when added one at a time. The interaction between cognateness and En-

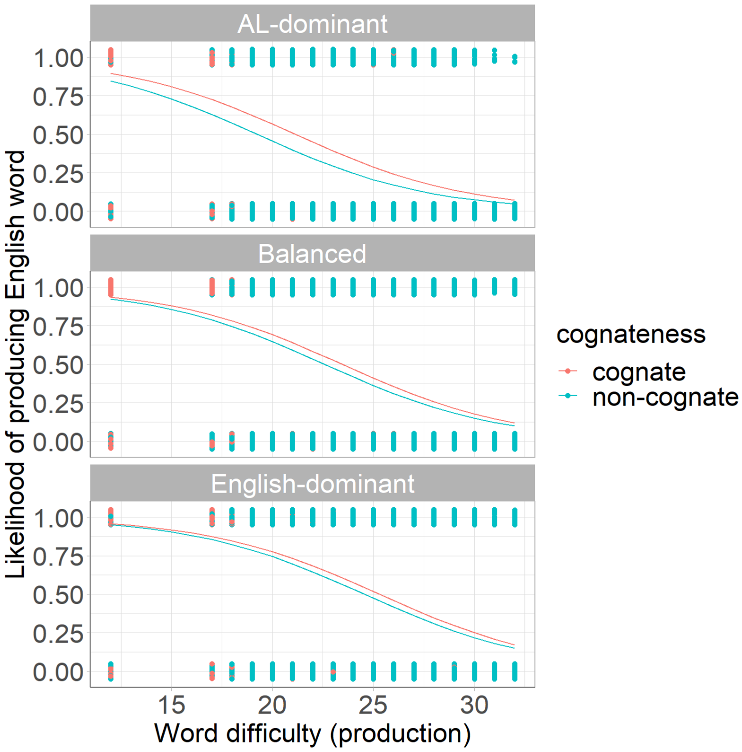

Figure 5.3: Plot visualising model predictions for the generalised linear model of English production, with lines of best fit showing the probability of each word being produced depending on its word difficulty. The plots are split by cognateness and faceted by English exposure.

glish exposure did not significantly improve the model as judged using the reduction in marginal R^2 between model P3 and model P4.

Table 5.5: Model comparisons for English production. Change statistics compare each model with the one directly above it.

Model	R^2m	R^2m change	AIC	Chisq	Pr(>Chisq)
P0	.352		121181		
P1	.389	.0368	117166	4016.7912	$<$.001
P2	.390	.000763	117105	62.2296	$<$.001
P3	.392	.00187	116881	226.8671	$<$.001
P4	.392	-.000109	116877	5.7917	.0161

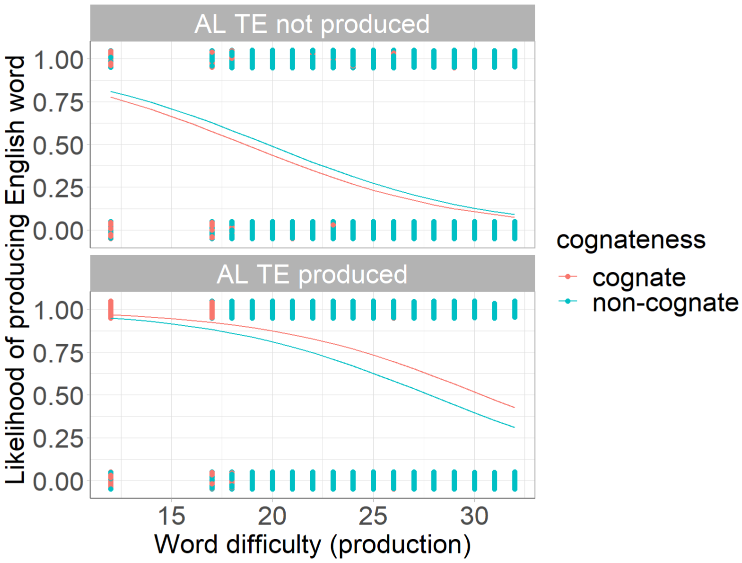

Figure 5.4: Plot visualising model predictions for the generalised linear model of English production, with lines of best fit showing the probability of each word being produced depending on its word difficulty. The plots are split by cognateness and faceted by whether or not the child produces the AL TE.

Table 5.6: Model coefficients for the full generalised linear mixed effect model of English production (Model P4), with age in months, English exposure, AL TE knowledge, word difficulty and cognateness as predictors, an interaction between AL TE knowledge and cognateness, and an interaction between English exposure and cognateness.

Predictor	Estimate	Std Error	z	p
(Intercept)	-1.203	0.090	-13.369	$<.001$
Age	1.263	0.080	15.733	$<.001$
Word difficulty	-1.199	0.036	-33.191	$<.001$
English exposure	0.893	0.083	10.772	$<.001$
AL TE production	1.141	0.021	55.197	$<.001$
Cognateness	-0.039	0.044	-0.883	.3774
Cognateness : AL TE prod.	0.779	0.057	13.7	$<.001$
Cognateness : English exposure	-0.072	0.030	-2.423	.0154

5.2.3 Study 1 Discussion

Study 1 provided supported that toddlers were more likely to understand and to produce cognate words over non-cognate words after controlling for the word difficulty. The cognate facilitation effect was stronger for words where the toddlers also knew the translation equivalent. The cognate facilitation effect on English vocabulary was also stronger in toddlers for whom English was their less-dominant language, but the effect could be attributed to existing knowledge of words in their dominant language facilitating the acquisition of cognates in their L2 (English).

One limitation for the use of cross-sectional data to study the effect of existing word knowledge is that we cannot investigate differentiate whether the two cognate translation equivalents are acquired in parallel or one after the other. By using analyses of longitudinal vocabulary growth, we will be able to study the acquisition of translation equivalents over consecutive time points. I will analyse word pairs that were either (1) understood/produced as a singlet at time point 1, or (2) not understood/produced at time point 1. For example, if a child knows the English word “fish” but not its Dutch translation “vis”, they are considered to know the word pair as a singlet. If the child knows neither “fish” nor “vis”, they are considered not to know the word pair. I investigate the likelihood that a word will be learnt by time point 2, dependent on its cognate status and whether or not the translation equivalent was known at time point 1. Under my hypothesis that cognate facilitation relies on prior knowledge, we would expect that if a child knows the English words “fish” and “duck” at timepoint 1, they will be more likely to have learnt the Dutch word “vis” (fish; cognate) than “eend” (duck; non-cognate) at timepoint 2. In contrast, we would expect the child to be equally likely to learn “vis” and “eend” by timepoint 2 if they did not know “fish” nor “duck” at timepoint 1. The longitudinal analyses in Study 2 will act as confirmatory analyses to my hypothesis that cognate facilitation occurs only when the learner has is existing knowledge of the cognate word's translation equivalent.

5.3 Study 2: Longitudinal

5.3.1 Methods

Participants

The sample of Study 2 was the subset of 210 families from Study 1 who contributed longitudinal data. Longitudinal families contributed between 2 to 7 questionnaires in total, with the total number of questionnaires totalling 616. The questionnaires were sent to parents at 2-month intervals while their child was between 12 and 32 months-old, but as parents completed the questionnaire in their own time and may not answer every questionnaire sent to them, the interval between consecutive longitudinal questionnaire responses varied between 38 days and 443 days. The mean interval between timepoints was 104.5 days (SD = 57.7).

Analysis Plan

The longitudinal data was analysed to further evaluate the findings of Study 1. I defined each timepoint as T(n) and the timepoint immediately following it as T(n+1). I analysed word pairs that were known at T(n) in only one language (*Tn-singlet*) or not known in either language (*Tn-unknown*). Trials where the toddler already knew both words in a pair at T(n) were excluded from the analysis.

Word knowledge at T(n) was included as a main predictor. I also included a main predictor of cognateness, and an interaction between cognateness and word knowledge at T(n). The dependent variable was whether or not a previously-unknown word in the pair was learnt by T(n+1). This is defined as acquiring an additional word in a translation equivalent pair. For pairs where one of the pair was known at T(n), the pair should be known as a doublet at T(n+1) to be classified as learnt. For pairs where neither word was known at T(n), the pair should become known as a single or a doublet at T(n+1) to be classified as learnt.

The model syntax was specified as below in R:

```
glmer(learnt_by_T(n+1) ~ cognateness +
    T(n)_word_knowledge +
    T(n)_word_knowledge:cognateness +
    (1|participant/timepoint) + (1|concept))
```

The reference level for T(n) word knowledge was Tn-unknown. The reference level for cognateness was non-cognates. I predicted that, following the pattern observed in Study 1, cognates would be more likely to be learnt at T(n+1) than non-cognates, but only if one of the words was already known at T(n). Statistically, this would manifest as a significant positive interaction effect in the model.

5.3.2 Results

Comprehension

A total of 71844 trials were analysed for comprehension. Of these trials, 46411 were Tn-unknown non-cognates, 3808 were Tn-unknown cognates, 20194 were Tn-singlet non-cognates, and 1431 were Tn-singlet cognates. An additional 3038 trials were excluded from the analysis for showing a decrease in word knowledge over consecutive time points, i.e. being reported as a singlet at T(n) but as an unknown word at T(n+1). These trials were excluded as cases of reporting error. Small percentages of reporting error are expected in parent-report vocabulary measures.

As predicted, cognates were more likely to show increase in word knowledge between consecutive timepoints than non-cognates, with the effect of cognates stronger when at least one word was already understood at T(n). This pattern of results is visualised in Figure 5.5.

When neither translation equivalent was understood at T(n), toddlers were equally likely to learn one or more translation equivalents of the pair regardless of whether the words were cognates or non-cognates. To give an example, a Dutch-English child who knew neither “fish” nor “duck” at T(n) was equally likely to learn “vis” (fish) and “eend” (duck) by T(n+1). In contrast, there was a striking difference for words which were singlets at T(n), i.e. when 1 of the pair was already known at T(n). At T(n+1), the previously-unknown translation equivalent was more likely to be learnt when the pair were cognates than when they were non-cognates. A

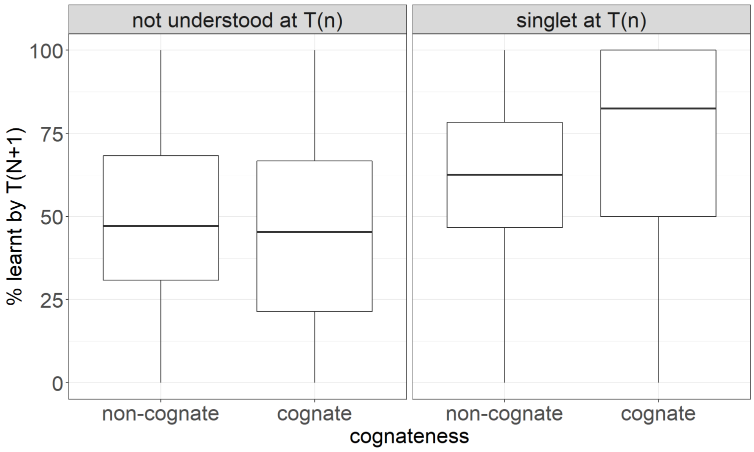

Figure 5.5: Boxplots showing the percentage of cognates and non-cognates that showed increase in word comprehension between T(n) and T(n+1), split by whether the word was understood in one language at T(n) or not understood in either language. Percentages were calculated for each participant.

child who knew both “fish” and “duck” at T(n) was more likely to have learnt “vis” (fish; cognate) than “eend” (duck; non-cognate) by T(n+1).

This observed pattern is supported by model coefficients shown in Table 5.7. The significant positive interaction of word knowledge at T(n) and the word’s cognateness indicates that the effect of cognateness is larger for words known as singlets at T(n) than words not known at T(n). Interestingly, the main effect of cognateness was not significant, suggesting that cognates were not easier to learn than non-cognates overall. Taken together, the model suggests that existing knowledge of a translation equivalent is necessary to support the easier acquisition of cognate words over non-cognate words.

Notably, there was a *negative* main effect of T(n) word knowledge, where toddlers where less likely to acquire an additional translation equivalent when they already knew one word for the pair as compared to a new concept. This is incompatible with the theory that toddlers have a preference for learning translation equivalents as discussed in Chapter 2.

Table 5.7: Model coefficients for the generalised linear mixed effect model for comprehension at (Tn+1), with T(n) word comprehension and cognateness as predictors, and an interaction between T(n) word knowledge and cognateness.

Predictor	Estimate	Std Error	z	p
(Intercept)	0.694	0.119	5.85	<.001
Cognateness	-0.0236	0.0549	-0.431	.667
T(n) word comprehension	-0.433	0.0248	-17.4	<.001
Cognateness : Tn word knowledge	0.546	0.0813	6.72	<.001

Production

A total of 108369 trials were analysed for production. Of these trials, 85583 were Tn-unknown non-cognates, 7089 were Tn-unknown cognates, 14518 were Tn-singlet non-cognates, and 1179 were Tn-singlet cognates.

Vocabulary growth in production followed a similar pattern as comprehension, extending results found in Study 1. Again, cognates were more likely to show increase in word production between consecutive timepoints than non-cognates, with the effect of cognates stronger when at least one word was already produced at T(n). This pattern of results is visualised in Figure 5.6.

As with comprehension, there was no difference in the likelihood of a word being learnt at T(n+1) between cognates and non-cognates when neither translation equivalent was produced at T(n). The advantage of cognates over non-cognates was only seen when a singlet was produced at T(n). A child who produced both

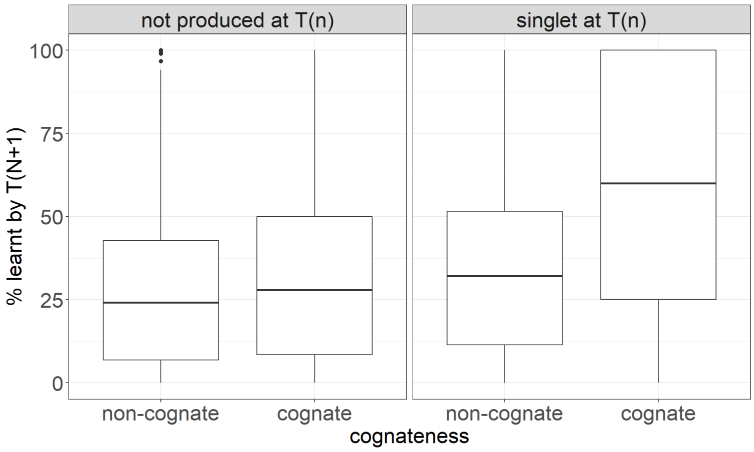

Figure 5.6: Boxplots showing the percentage of cognates and non-cognates that showed increase in word production between T(n) and T(n+1), split by whether the word was produced in one language at T(n) or not produced in either language. Percentages were calculated for each participant.

"fish" and "duck" at T(n) were subsequently more likely to also have learnt to produce "vis" (fish; cognate) than "eend" (duck; non-cognate) at T(n+1). Again, we see in the model coefficients (Table 5.8) that the interaction between T(n) word production and cognateness was significant and positive. This mean that the facilitatory effect of cognates on the production of new words was stronger when the child already produced one of the translation equivalent pair at T(n). The main effect of cognateness was positive for production, suggesting that cognates were easier to produce than non-cognates. Nevertheless, the significant interaction indicated that prior production of singlets further boosted the advantage of cognates over non-cognates.

The pattern observed in comprehension for T(n) word comprehension is reflected in the production model for T(n) production. The negative main effect suggests that toddlers are less likely to learn to produce the translation equivalent of a known word than they are to learn to produce a novel word. This unexpected pattern of results is discussed further in the discussion of this study.

Table 5.8: Model coefficients for the generalised linear mixed effect model for production at Tn+1, with Tn word production and cognateness as predictors, and an interaction between Tn word knowledge and cognateness.

Predictor	Estimate	Std Error	z	p
(Intercept)	-1.368	0.168	-8.12	$<$.001
Cognateness	0.260	0.0488	5.31	$<$.001
Tn word production	-1.525	0.0286	-53.36	$<$.001
Cognateness : Tn word knowledge	0.659	0.0817	8.07	$<$.001

5.3.3 Study 2 Discussion

The longitudinal analyses further supported my hypothesis that phonological similarity across translation equivalents facilitate the acquisition of a previously-unknown word, but only when there was existing knowledge of the translation equivalent in the other language. This advantage of cognates was observed for both comprehension and production.

Another notable pattern observed in Study 2 is that toddlers showed evidence of a preference towards learning *novel concepts* over translation equivalents, with the exception of cognates. This effect remained robust even after adding additional covariates of toddlers' age at the second time point and the time that elapsed between time points. This negative coefficient is in direct contrast to the patterns observed in Study 1 where it was more likely for an English word to be known if its AL translation equivalent was also known. The differing results may be attributable to the difference in analysis method. In Study 1, words known in one language were likely to be also known in the other. In contrast, the findings in Study 2 appear to provide evidence against the idea that toddlers preferentially learn translation equivalents. Instead, it would suggest that the high proportion of doublets observed in bilingual toddlers' vocabulary in Chapter 2 may be simply be an artifact of similar words being easiest to learn across the two languages. This is in direct contrast to the hypothesis I proposed at the beginning of this thesis, that toddlers would find it easier to map a new word form to a known concept rather than to learn a new concept.

Cognates are the exception to this observed pattern. There was a positive interaction effect of cognates and T(n) word comprehension despite the (non-significant) negative main effect of cognates and significant negative main effect of T(n) word comprehension. This suggests that despite toddlers having a preference to learn novel words and cognates being harder to learn, the scaffolding provided by having two semantically-equivalent word forms with high phonological similarity supports toddlers' vocabulary acquisition.

5.4 Study 3: Vocabulary size

The aim of Study 3 was to investigate whether the cognate facilitation effect observed at the word level in Studies 1 and 2 translates into a larger vocabulary size for children whose languages have more cognates, and particularly if it can also be associated with advantages in learning translation equivalents. I put forward two main hypotheses. Firstly, I predict that learners of languages with a higher percentage of cognates will have larger total vocabularies than learners of languages with less cognates. Total vocabulary size sums the respective vocabulary sizes in the bilingual's two languages as judged using a bilingual vocabulary questionnaire. Secondly, I predict that learners of languages with many cognates will have a higher proportion of doublets in their vocabulary than learners of languages with less cognates.

5.4.1 Methods

Participants

The sample of Study 3 was the same cross-sectional sample as in Study 1, described in Section 5.2.1.

Calculating vocabulary size

I derived two measures of vocabulary size from the CDI data for each toddler, to evaluate toddlers' vocabulary growth trajectories.

Total vocabulary size (Equation 5.1) was used as a measure of ease of vocabulary acquisition, where a larger vocabulary size is associated with greater learning ease.

$$Total\ vocabulary\ =\ English\ vocabulary + AL\ vocabulary \tag{5.1}$$

Proportion of doublets (Equation 5.2) was used to measure the degree of concep-

Table 5.9: Percentage of cognates in CDI between English and each AL

AL	language family	% cognates	% cognates (scaled)
Dutch	Germanic	20.3%	1.24
German	Germanic	16.0%	0.82
French	Romance	7.84%	0
Italian	Romance	7.84%	0
Spanish	Romance	7.84%	0
Portuguese	Romance	5.56%	-0.23
Polish	Slavic	4.25%	-0.36

tual overlap between a child's vocabulary in their two languages, where a higher proportion suggests that children find it easier to learn both translations. Doublets refer to translation equivalents where both the English word and the AL word are understood by the child. The number of doublets known is divided by the total concepts known (where the child understands the word in English, in the AL or in both languages) to obtain a percentage of doublets in their vocabulary.

$$Proportion\ of\ doublets = \frac{no.\ of\ doublets}{conceptual\ vocabulary} \tag{5.2}$$

Calculating percentage of cognates

I calculated the percentage of cognates between English and each AL among the 306 words overlapping across the CDIs. Cognates were defined as in Study 1. As mentioned in Study 1, Dutch and German shared the largest number of cognates with English by a large margin. This was followed in order by French, Italian, Spanish, Portuguese and Polish (Table 5.9). Percentage of cognates was centred on the median (7.84%) and scaled by intervals of 10% for the analysis.

5.4.2 Results

I expect that the total vocabulary size of toddlers learning language pairs with many cognates (e.g., English-Dutch) will be larger than that of children learning languages with few cognates (e.g., English-Polish). Additionally, I predict that the availability of cognates would make it easier for toddlers to learn translation equivalents, which should manifest as a higher proportion of doublets in the vocabularies of toddlers learning languages with more cognates. I expect that a toddler learning English and Dutch would know more words and have more doublets in their vocabulary than a toddler learning English and Polish. To test these predictions, I ran two mixed effects models using the `lme4` package (Bates, Mächler, Bolker, & Walker, 2015), the first with total vocabulary size as the dependent variable and the second with percentage of doublets. Both models had cognate percentage (centred on median and scaled by intervals of 10%) as the main predictor, with child's age (centred on mean age and scaled by SD) and overall English exposure (centred on 50% and scaled by SD) as covariates. Overall English exposure was added as a covariate because there is imbalance in the distribution of this statistic between languages. Language pair was added as a random effect. The model syntax is shown below:

```
lmer(total_vocabulary_size ~ age +
    overall_English_exposure +
    percentage_of_doublets + (1|language_pair))
```

P-values were obtained using lmerTest (Kuznetsova et al., 2017) via Satterthwaite's approximation degrees of freedom. Median cognate percentage was 11%

and and a 1 unit change in scaled cognate percentage represents 10% more cognates across the language pair.

Comprehension (12-26 months old)

As predicted, I found a significant positive main effect of cognate percentage on total vocabulary size in comprehension (Table 5.10; Figure 5.7). The statistics show that languages pairs with 10% more cognates are associated with an average total vocabulary size in comprehension that is 2.73% larger. I likewise found a significant positive main effect of cognate percentage on the percentage of doublets (Table 5.11; Figure 5.8). This means that children learning languages pairs with 10% more cognates understand on average 2.94% more doublets.

Table 5.10: Mixed effect model for total vocabulary size in comprehension, with age, English exposure and percentage of cognates as predictors, and language pair as a random effect.

Predictor	Estimate	Std Error	t	p
(Intercept)	41.02	0.962	42.6	$<$.001
Age	18.61	0.936	19.9	$<$.001
English exposure	0.252	0.940	0.268	.789
% cognates	5.91	2.164	2.73	.00657

Table 5.11: Mixed effect model for percentage of doublets in comprehension, with age, English exposure and percentage of cognates as predictors, with language pair as a random effect.

Predictor	Estimate	Std Error	t	p
(Intercept)	45.67	1.01	45.1	$<$.001
Age	11.73	0.985	11.9	$<$.001
English exposure	2.33	0.989	2.36	.019
% cognates	6.69	2.28	2.94	.0035

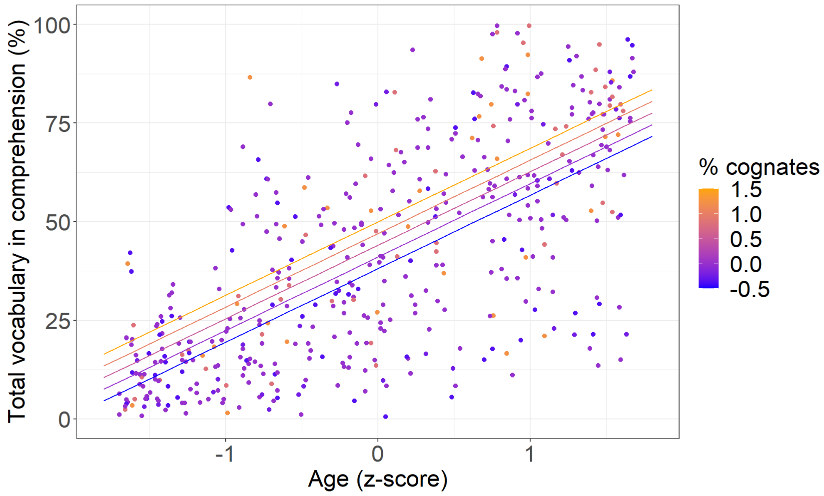

Figure 5.7: Model predictions for total vocabulary size in comprehension with the main effect of language similarity (defined by log-transformed percentage of cognates) as lines of best fit, and with child's age on the x-axis.

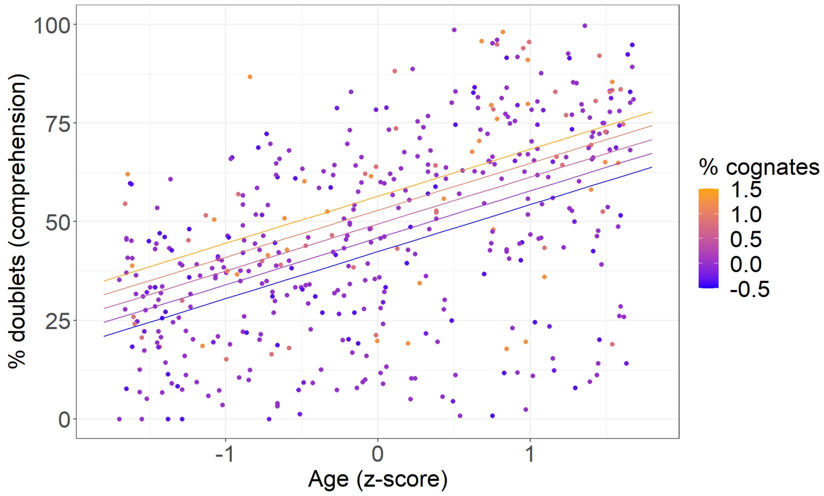

Figure 5.8: Model predictions for percentage of doublets in child's vocabulary in comprehension with the main effect of language similarity (defined by log-transformed percentage of cognates) as lines of best fit, and with child's age on the x-axis.

Production (20-36 months old)

The pattern observed in production was very similar to the pattern observed in comprehension. As predicted, I found a significant positive main effect of cognate percentage on total vocabulary size in production (Table 5.12) (Figure 5.9), mirroring the pattern observed in comprehension. The statistics show that languages pairs with 10% more cognates are associated with an average total vocabulary size that is 3.74% larger. I likewise found a significant positive main effect of cognate percentage on the percentage of doublets (Table 5.13) (Figure 5.10). This means that children learning languages pairs with 10% more cognates produce on average 5.77% more doublets.

Table 5.12: Mixed effect model for total vocabulary size in production, with age, English exposure and percentage of cognates as predictors, and language pair as a random effect.

Predictor	Estimate	Std Error	t	p
(Intercept)	37.9	0.967	39.2	$<$.001
Age	15.1	0.893	16.9	$<$.001
English exposure	1.52	0.919	1.65	.0995
% cognates	7.03	1.88	3.74	.000204

Table 5.13: Mixed effect model for percentage of doublets in production, with age, English exposure and percentage of cognates as predictors, with language pair as a random effect.

Predictor	Estimate	Std Error	t	p
(Intercept)	34.5	1.016	33.9	$<$.001
Age	7.64	0.938	8.14	$<$.001
English exposure	-0.236	0.965	-0.244	.807
% cognates	11.4	1.97	5.77	$<$.001

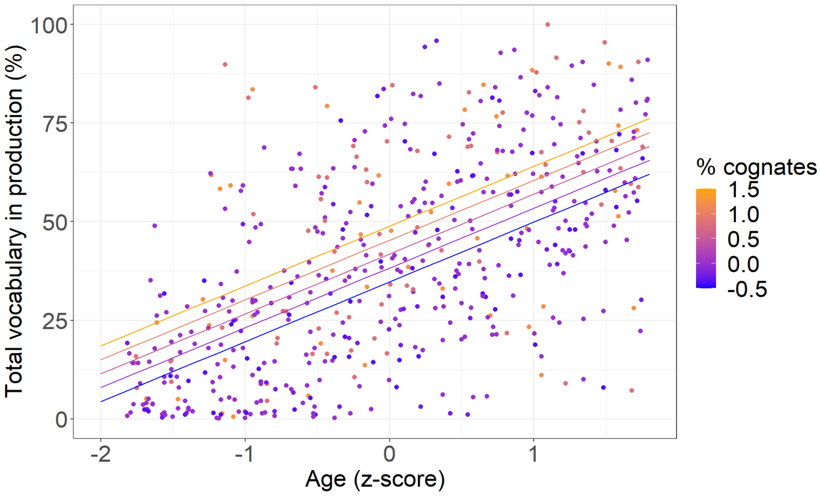

Figure 5.9: Model predictions for total vocabulary size in production with the main effect of language similarity (defined by log-transformed percentage of cognates) as lines of best fit, and with child's age on the x-axis.

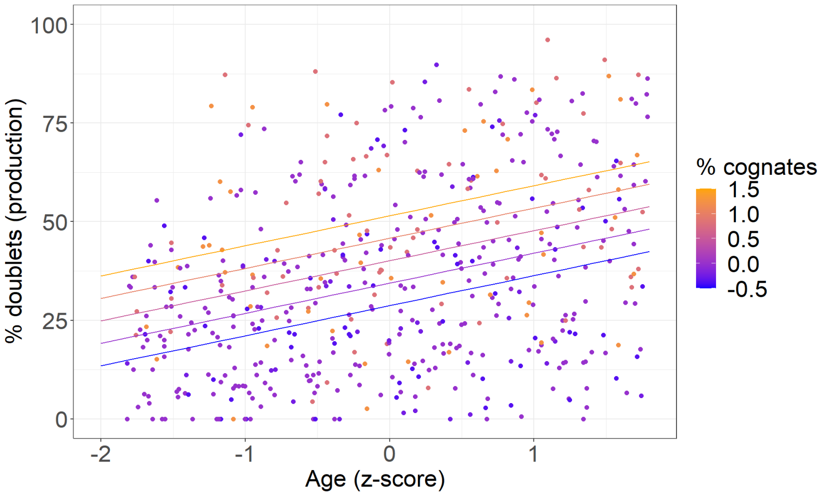

Figure 5.10: Model predictions for percentage of doublets in child's vocabulary in production with the main effect of language similarity (defined by log-transformed percentage of cognates) as lines of best fit, and with child's age on the x-axis.

5.4.3 Study 3 Discussion

In Study 3, I found that the cognate facilitation effect observed in Studies 1 and 2 translates into larger vocabulary sizes in bilingual toddlers whose language pair has more cognates. Additionally, toddlers whose languages have more cognates also have proportionally more doublets in their vocabulary.

Findings in this study are consistent with that of Floccia et al. (2018), a large scale CDI study testing the effect of language distance conducted with bilingual toddlers in the UK. Floccia et al. collected vocabulary data from 24-month-old bilingual toddlers growing up with English and one additional language. They used the average Levenshtein distance of 406 translation equivalents in the CDI to quantify the phonological distance between language pairs. They found a significant relationship between phonological distance and toddlers' vocabulary in production. Toddlers who were learning phonologically closer language pairs had higher production vocabulary in their AL than those learning more dissimilar languages. Vocabulary in comprehension was linked to two other metrics of language similarity – typological similarity (i.e., whether the languages shared the same word order for verbs and objects in a sentence) and morphological similarity (i.e., whether the languages were similar in the complexity of morphological markers used to indicate number, tense, person and other grammatical aspects). Toddlers whose languages were typologically closer or morphologically closer had larger comprehension vocabulary. In Study 3 in this chapter, I extended the research direction of Floccia et al. (2018) in investigating the links between phonological distance and vocabulary size with a larger word list. In Floccia et al. (2018), English vocabulary size was collected using a short-form of the Oxford CDI, which only had 30 words that overlapped across all the CDIs used to measure vocabulary knowledge in toddlers' ALs. In the present study, I had 306 word pairs in total that overlapped across all the tested languages. This resulted in improved statistical power for the analyses.

There was a difference in the method used to define language distance between the present study and Floccia et al. (2018). Instead of using the average phonological distance of all words to define phonological distance, I calculated the proportion of cognates. This was chosen following my hypothesis that cognates facilitate bilingual word learning by having obvious word-level phonological similarity. When using an average phonological distance metric, a language pair where all word pairs have moderate similarity may receive a similar score to a language pair where half the word pairs have high similarity and half the word pairs have low similarity. By using the proportion of cognates to define language similarity, I aimed to identify languages that contain words with high word-level similarity.

5.5 Discussion

5.5.1 Cognate facilitation in toddlers' vocabulary

In this chapter, I presented findings that support the facilitatory effect of cross-linguistic word similarity on bilingual toddlers' vocabulary trajectories. Results from the cross-sectional sample in Study 1 are further supported by analyses on longitudinal data in Study 2. Across both Study 1 and Study 2, cognates were more likely to be known than non-cognates, but only if the toddler also knew the translation equivalent in their other language. This effect of cognates was seen both in comprehension and production. This context-dependent cognate facilitation effect is striking, as it supports hypotheses that the cognate facilitation effect manifests when learners are able to compare novel words with existing lexical representations. By identifying similarities between novel words and the words they already know, learners are able to make use of those similarities to learn novel words more easily. These results complement existing literature on cognate facilitation in second language acquisition. I show that even toddlers are sensitive to cross-linguistic similarity at the word level.

Additionally, any effects of language dominance on cognate facilitation in vocabulary trajectories appears to be attributable to whether or not the toddler also knows the translation equivalent in the other language. For both English comprehension and production, including an additional interaction between English exposure and cognateness did not increase the marginal R^2 of the model compared to a simpler model that already included the interaction between AL TE knowledge and cognateness. I take this to suggest that the cognate facilitation effect in L2 vocabulary learning is dependent on existing knowledge of the translation equivalent in learners' L1. An AL-dominant bilingual toddler who knows more words in the AL than in English may find it easier to learn cognates than non-cognates in English, but only if the AL translation equivalent is known. At the same time, even an English-dominant bilingual could benefit from cognateness for a word they have not yet learnt in English if they know the AL translation equivalent.

5.5.2 Preference for translation equivalents or new concepts

Another notable result of Study 1 is the unique variance explained by AL TE knowledge for both comprehension and production, even after accounting for the co-variates of age and word difficulty. I interpret this result to mean that bilingual toddlers have a preference to learn translation equivalents of concepts they have acquired. One hypotheses behind this preference can be linked to the cognitive demands of learning new concepts. When learning a new word, learners have to not only familiarise themselves with the word form (phonology or orthography), but also its meaning. Learning the meaning of the concept ’dog’ involves both learning the features that identify a dog, and also learning how to differentiate a dog from other animals. The acquisition of concepts is a complicated task, even before involving the phonological features of the word form. For learners of two languages, translation equivalents share their meaning, allowing two word forms to be mapped onto the same concept node. Therefore, I theorise that mapping a new word form to an existing concept would be cognitively easier than learning a new concept. This lower cognitive demand may explain a preference for learning translation equivalents. However, contradictory findings were found in Study 2, with the pattern of results suggesting that new concepts might be preferred over translation equivalents instead. This observed pattern was unexpected and warrants further research and replication to check if the results are reliable.

The sample in the present study was recruited from countries with predominantly monolingual communities (UK, Netherlands, Germany, Spain). The demographics are thus sociolinguistically different from the communities studied by some bilingual researchers, for example Spanish-Catalan in Barcelona (Bosch & Ramon-Casas, 2014). Instead, the sample is comparable to that of Floccia et al. (2018), being made up of bilingual infants growing up with early input to one language widely spoken in the community and a heritage language spoken only by subgroups in the community. A feature of heritage bilinguals is that the language input received outside the home will predominantly be the community language (in the case of the UK sample, English). Conversely, the AL will be mostly heard inside the home. This creates an imbalance in the quantity of language exposure and the variety of contexts each language is heard in. The two languages may thus grow at different rates and there may be unique challenges to the acquisition and maintenance of each language. Language requirements of different communicative partners (who may not speak one of the languages) may prompt bilingual toddlers to preferentially learn translation equivalents, thus acquiring the same concepts in each languages to allow effective communication. Alternatively, if the two languages are largely used in different contexts, toddlers may preferentially learn new concepts to fill in gaps in their conceptual vocabulary. Comparisons between the ratio of translation equivalents in early vocabulary of bilinguals growing up in predominantly monolingual communities and that of bilinguals from bilingual communities would shed light on the influences of the learning environment on bilingual vocabulary trajectories.

5.5.3 Cognates and vocabulary size

In Study 3, I found that toddlers learning languages that have more cognates had larger vocabulary sizes and knew more doublets than toddlers learning languages with less cognates, even after controlling for the child's age. These findings support the idea that languages with more cognates are easier to learn than those with less cognates. Taken together with findings from Studies 1 and 2 that cognates are more likely to be known by toddlers than non-cognates, I theorise that phonological similarity between translation equivalents helps learners to disambiguate the meaning of words more easily, facilitating the process of learning two languages in parallel. The results complement existing literature on the relationship between cross-linguistic similarity and bilingual toddlers' vocabulary size (Floccia et al., 2018; Gampe et al., 2021).

5.6 Conclusion

In this chapter, I provide evidence suggesting that toddlers can capitalise on the strong phonological overlap between cognates to facilitate learning of translation equivalents, thus expanding their vocabulary. Notably, the cognate facilitation effect is dependent on existing knowledge of the word's translation equivalent in the other language. This provides support to the idea that even toddlers have the ability to compare and contrast their languages. These bilingual toddlers do not learn their two languages in isolation but instead actively make use of both semantic and phonological overlap between words in their languages to facilitate this difficult process. The shared properties of the languages being learnt can help toddlers learn their languages more easily, subsequently resulting in larger vocabulary sizes and more doublets in vocabulary for toddlers whose languages are more similar. Understanding the links between languages and the effect on bilingual vocabulary acquisition can help guide strategies for supporting bilingual vocabulary.

Chapter 6

Exploring the variable effects of frequency and semantic diversity as predictors for a word's ease of acquisition in different word classes

6.1 Background

Children learn language via experience and exposure. The language environment of a child can include child-directed speech from their parents (Hart & Risley,

---

Study 1 in this chapter was published with the following reference: Siow, S., & Plunkett, K. (2021). Exploring the variable effects of frequency and semantic diversity as predictors for a word's ease of acquisition in different word classes. Proceedings of the Annual Meeting of the Cognitive Science Society, 43. Retrieved from https://escholarship.org/uc/item/83t6n1rq. SS was responsible for data analysis, interpreting results and writing the manuscript.

1992; Rowe, 2012; Warren-Leubecker & Bohannon III, 1984), child-directed speech from other children (Weisner et al., 1977; Cristia, 2022), and overheard speech directed by adults towards each other (Shneidman & Goldin-Meadow, 2012) or by adults towards the child's older siblings (Oshima-Takane, Goodz, & Derevensky, 1996). The proportion of child-directed speech and overheard speech differs between cultures, with Western industrialised cultures tending to have more focus on parent-child dyads while it is common in many other cultures to involve additional family members and community in child rearing (Barry & Paxson, 1971; Weisner et al., 1977). It can also vary depending on family's socioeconomic status and parents' knowledge of child development (Rowe, 2008). For bilinguals, child-directed speech can differ from overheard speech in the the proportion of the two languages used (Orena et al., 2020). Children have been shown to be able to learn from overheard speech (Akhtar, Jipson, & Callanan, 2001) as well as child-directed speech (Foursha-Stevenson, Schembri, Nicoladis, & Eriksen, 2017).

The language environment analysed in the present study will focus on child-directed speech from adults, with data obtained through transcripts of parent-child interactions available on CHILDES. Analyses of child-generated productions reveal that the amount and variety of words used by children correlates highly with the words used by their mothers (H. Li & Fang, 2011). Both child-directed speech and child-produced speech are predominantly composed of concrete nouns and common pronouns, which have obvious referents. Hills (2013) found that child-directed speech possesses higher rates of word repetitions, less diversity and more semantic associations between adjacent words than adult-directed speech.

6.1.1 Quantifying word acquisition

A common methodology for measuring infant word knowledge is via parental-report vocabulary questionnaires. At the group level, a word's *Age of Acquisition* (AoA) is typically operationalised as the youngest age at which more than 50% of sampled children are reported to understand the word (in the case of AoA measured in comprehension) or produce it (AoA in production). This value of 50% is in truth arbitrary, however this simple threshold is sufficient for the purposes of this study. For the analyses explored in this chapter (and in fact much of the literature), we are less concerned about the exact value of AoA, and more about the *order* of acquisition. The quantified statistic of AoA is commonly used as a proxy of a word's ease of acquisition relative to other words. In all further discussions of AoA in this chapter, I will be referring to this measure of relative ease of acquisition.

6.1.2 Frequency and AoA

What factors make a word easier to learn? Children have to deal with complexities in both the linguistic and physical environment. One feature of this complexity is that children are not necessarily oriented towards the correct referent of the word at the time of naming. Naming events are just as likely to be unambiguous (with the referent in a dominant visual position) as highly ambiguous (with the referent not present or unclear) from the child's visual perspective (Yurovsky, Smith, & Yu, 2013). Repetitions and high frequency of words may make it easier to map the correct referent to the object, because they increase the frequency of the word co-occurring with its referent, and particularly occurrences where the child's attention is on the referent. Many studies have found correlations between a word's AoA and its frequency of occurrence in the language environment (Goodman et al., 2008; Hills et al., 2010; Hills, 2013; Braginsky et al., 2019).

Quantifying frequency in CDS

Transcripts of Child Directed Speech (CDS) can be used to obtain quantitative measures of children's early learning environments. A word's frequency is quantified by counting the total number of times that word appears in the corpus. Frequency may be counted for word forms (i.e., separate counts for "run", "ran", "running", etc.) or for lemmas (i.e., combining all the above instances under the concept "run"). It is important to note that frequency as derived from corpus data can only give us direct data on word form occurrences. However, word learning doesn't only involve learning a word form, but also the mapping of the word form to its referent. Studies of frequency using corpus data make the implicit assumption that words are typically uttered in the presence of their referent, allowing for word-referent mapping. Given the extensive literature on contingent responding and joint attention in caregiver-child interactions, this assumption seems a reasonable one. Parents preferentially name and refer to objects that their child is looking at. Joint attention on the referent by both parent and child during word utterances has been found to be predictive of successful learning (Tomasello & Todd, 1983).

6.1.3 Contextual diversity and AoA

The hypothesised effect of *contextual diversity* on word learning is rooted in the cross-situational learning literature (Pinker, 1994; Gleitman, 1990). Infants are able to extract statistical information about a word-referent pairing by rapidly compiling information across multiple occurrences and evaluating the regularities within the input (L. Smith & Yu, 2008). Several explanations have been proposed regarding the mechanism(s) underlying cross-situational learning. One possibility is that learners operate using associative learning mechanisms (McMurray, Horst, & Samuelson, 2012). Co-occurrences between word and referent may facilitate resolution of referential ambiguity. Alternatively, learners may engage in Bayesian hypothesis testing, where learners assign probabilities to each object in an ambiguous scene, representing the likelihood that a given object is the target word's referent (Xu & Tenenbaum, 2007). These probabilities are derived and adjusted using evidence over multiple scenes. Probabilities can be increased through systematic occurrences of a word-object pair, and decreased by the absence of a given possible referent during utterance of the word. The object with the highest probability can be inferred to be the intended referent, subject to collecting sufficient evidence.

Given the same number of occurrences, word-referent pairs are learnt more successfully when a target word-referent pair co-occurs in a wide variety of contexts than when it appears with only one or two unique word-referent pairs (Kachergis, Shiffrin, & Yu, 2009; Suanda, Mugwanya, & Namy, 2014). By occurring in a variety of contexts in the joint presence of different objects, word-object pairs can be assigned different weightings (whether by associative learning or hypothesis testing) depending on systematic co-occurrences or non-co-occurrences, thus facilitating inference of word meaning.

Quantifying contextual diversity in CDS

Besides frequency, corpora of CDS transcripts can also be used to identify the words that a given word co-occurs with. The number of unique words a given word co-occurs with in CDS can be used to quantify the diversity of the linguistic environment (and by proxy the physical environment) that a given word occurs in. Hills et al. (2010) referred to this quantified variable as contextual diversity, forming a link with the cross-situational learning literature. As with corpus-derived frequency, this measure of contextual diversity also relies on several assumptions: firstly, that words occur in the presence of their referents, and secondly that the diversity of word co-occurrences corresponds reliably with variation in the physical environment, so as to give cues for inference of meaning. Hills et al. tested the predictor of contextual diversity in child-directed speech for predicting AoA in toddlers' vocabulary. When quantifying contextual diversity from speech transcriptions, the definition of an edge could be as limited as only words directly adjacent to the node word, or include all words within a multi-word window. Hills et al. found that a window of 5 words was optimal for defining contextual diversity in child-directed speech to predict individual words' AoA. This measure of contextual diversity should include mostly syntagmatic word associations denoting sentence structure (e.g. "apple" – "eat"), though it can also include some paradigmatic associations of words belonging to the same category (e.g. "apple" – "pear"), as found by Wettler, Rapp, and Sedlmeier (2005) for adult speech. Chang and Deák (2020) showed that co-occurrences in child-directed speech include both syntactic and thematic relationships, and that both types of co-occurrence contribute unique variance to predicting AoA even after accounting for word frequency.

6.1.4 Semantic associations and AoA

Free associations have been used as a measure for semantic association strength between word pairs and also to quantify semantic richness surrounding a word. Adult-generated associative norms have been found to correspond with contiguity in adult speech (Wettler et al., 2005). Adult-generated associations were found to be predictive of words' AoA in infants (Hills et al., 2010, 2009). While these statistics have been widely supported as good predictors of performance in adult lexical decision tasks, much less it known how they might relate to AoA. Hills et al. (2010) proposed that the relationship between associations and AoA of early-learnt words may be partially accounted for by contextual diversity in everyday language.

When attempting to unpack what association norms may represent in an infant word acquisition context, we need to consider the sampling methodology that word association databases are built on. Unlike speech transcripts which record co-occurrences in everyday speech, word association data is collected experimentally by asking participants to list the first word(s) that come to mind. The South Florida Association Norms (Nelson et al., 2004) used by Hills et al. (2010) is an example of data collected using a discrete word association task, where participants could only give one response to each cue. Discrete word association tasks have been linked to more reliable indices of association strength and set size, but under-representation of weaker associates as compared to tasks that allow participants to give more than one answer (Nelson, McEvoy, & Dennis, 2000). Importantly, word associations have directionality that reflects their associative relationship and ease of retrieval. The cue "turtle" may elicit the response "animal", but the cue "animal" may preferentially elicit more frequent words like "dog" from within the large set of competitors within the animal category. For associative networks, *degree* includes all edges connected to the target word regardless of whether it was given as a cue or response. *in-degree* only counts the instances when the target word was given as a response, while *out-degree* only counts the instances when the target word was given as the cue. For example, if the cue "turtle" elicits the response of "animal" but the cue "animal" does not elicit the response of "turtle", the edge between "animal" and "turtle" would be included in the degree and out-degree of "turtle" but not the in-degree. As only words that are strongly related to the cue are likely to be produced as responses in a free association task, in-degree represents a subset of edges with comparatively strong relationships to the target word. Hills et al. (2009) found that associative in-degree was a better predictor of AoA than out-degree or overall degree.

Hills et al. proposed the *preferential acquisition model* for infant word acquisition, where words that are the most well-connected to other words in the learning environment via shared semantic relationships are most easily learnt. They suggested that this effect may stem from well-connected words being more salient within the learning environment and also that the richness of shared semantic context with related words can help inference of meaning. This is contrasted with the preferential attachment model that is frequently used to describe the growth of the lexicon in adults. The preferential attachment model posits that words that are connected to a larger number of words in the existing network are more likely to be learnt than words with less connections to the existing network. The preferential acquisition model emphasises connectivity to the language input the learner receives, while the preferential attachment model emphasises connectivity to the words the learner already knows.

6.2 Study 1: Predicting age of acquisition

The acquisition of language is sensitive to the linguistic environment that the learner is immersed in. Some words are naturally learnt earlier than others. This could be related to a variety of factors including the length of the word form, complexity of the concept, frequency of occurrence in speech, and availability of cues to infer its meaning. Study 1 in this chapter aims to investigate the word statistics of word frequency, contextual diversity in CDS and associative in-degree as predictors for AoA, extending Hills et al. (2010)'s study with British English data.

Frequency is expected to be a good predictor for the acquisition of all word classes, following the extensive literature linking word frequency and learning across several different word classes (Kachergis et al., 2009; L. R. Naigles & Hoff-Ginsberg, 1998; Hochmann, Endress, & Mehler, 2010). High word frequency is predicted to facilitate word acquisition through repeated exposures, with high frequency words having earlier AoA.

The other two predictors, contextual diversity and associative in-degree, quantify the richness of the learning environment that a word occurs in, proposed to support resolution of referential ambiguity. Words that have high contextual diversity and high associative in-degree are predicted to be learnt earlier as the rich context should provide toddlers with more cues for inferring the word's meaning and usage. The justification of including both predictors lies in composition differences as a result of their respective data sources. Contextual diversity, derived from co-occurrences in natural speech transcriptions, represents mostly syntagmatic relationships between words (e.g. subject "bunny", verb "eat" and object "carrot"). As such, contextual diversity holds information about the usage of a word in a sentence. It would be easier to interpret the meaning of the action "eat" if it occurs with a variety of food items (e.g. "carrot", "bread", "biscuit") as the object. The interpretation of "eat" is also supported by co-occurrence with animate subjects ("bunny", "child", "dog"). In comparison, associative in-degree, derived from free associations, represents mostly paradigmatic relationships (e.g. food items "carrot" and "biscuit"). Paradigmatic relationships helps learners to categorise concepts into groups. In the above example, both "carrot" and "biscuit" are both food items. At the same time, differences in the edges linked to words help learners to differentiate them. The vegetable "carrot" is more closely related to "lettuce" while the baked item "biscuit" is more closely related to "cake". As such, both syntagmatic relationships and paradigmatic relationships play a role to support interpretation of meaning. Given their differing compositions of syntagmatic and paradigmatic edges, it is valuable to include both contextual diversity and associative in-degree when analysing word acquisition order.

6.2.1 Methods

Age of acquisition

Full word-by-child comprehension and production scores were obtained from monolingual English data collected using the Oxford Communicative Development Inventory (CDI) (Hamilton et al., 2000). The Oxford CDI contains 418 words commonly known to young children. The inventory collects data via parental report for both comprehension and production of these words.

Three unpublished datasets were combined. The first dataset was collected between March 2020 and December 2020 using a 418-word Oxford CDI with random category presentation order (CDI2020), N = 180, age range 12–32 months old. The second dataset was made up of data collected between 2013 and 2020 using the 553-word extended version of the Oxford CDI (CDIExt), N = 330, age range 12–32 months old. In this version of the CDI, the 418 words of the standard-length CDI were always presented as the first 418 words in static presentation order. Analysis showed no significant difference between vocabulary scores collected with randomised-category presentation (CDI2020) or static presentation order (CDIExt) after controlling for the child's age (t = 0.267, p = .789). Participants in these two samples had taken part in lab-based studies at the Oxford University BabyLab. They completed the CDI as part of their participation in the studies. They were sent the link to an online questionnaire to complete the CDI at home within a week of their lab visit. They were offered a child-sized t-shirt as appreciation for their participation in the lab-based study. The CDI2020 sample is a subset of the monolingual sample reported in Chapters 2 and 5.

Finally, this data was combined with open-source data available on Wordbank (Frank et al., 2017), collected by the Plymouth BabyLab using the 418-word Oxford CDI, N = 1210, age range 12–25 months old. CDI data was divided into one-month chunks by child's age, representing completed months. Sample sizes in each age group ranged between 11 and 222. This data was used to calculate AoA in both comprehension and production for each word in the CDI. A word's AoA in comprehension was operationalised as the lowest age (in months) when the word reaches the threshold of being understood by at least 50% of toddlers at that age. Similarly, AoA in production was defined as the lowest age it is spoken by 50% of toddlers. There was a correlation of $.86$ between comprehension and production AoA (Pearson's r).

Four word classes were included in the analysis – nouns (N = 211), verbs (N = 65), adjectives (N = 36) and function words (N = 36). As in Hills et al. (2010), I excluded words about time (a very small class of 8 words), sounds, games and routines (an ambiguous word class, many of which are not single words), 8 words duplicated in multiple categories (e.g. noun "drink" and verb "drink") and an additional 6 words that are not single words (e.g. belly button).

Adult-generated associations

Adult-generated word pair associations were obtained from the Small World of Words (SWOW-EN) dataset (De Deyne et al., 2019), a database of free association responses for 12,292 cues. The SWOW-EN dataset was collected between 2011 and

2018. Out of the sample, the majority were adult native speakers of American English (81%), while 13% of participants reported British English as their first language. Participants were asked to type the first 3 words that come to mind in response to a given cue word. For comparability with the University of South Florida Free Association Norms (Nelson et al., 2004) used in Hills et al. (2010), I included only data corresponding to the first response given. *Associative in-degree* for a given word is defined as the total number of unique cues that elicit that word as a response. To avoid conflating this score with idiosyncratic responses, I only included words that were given by at least 2 participants for a given cue. This measure therefore represents the diversity of a word's strongest semantic associations. The associative in-degree of CDI words calculated from the whole sample had a $.96$ correlation with the British subset.

Corpora of naturalistic CDS

Corpora of CDS transcripts were downloaded from the CHILDES online database (MacWhinney, 2000) via the *childesr* package (Braginsky, Sanchez, & Yurovsky, 2020). Data was extracted from the British English subset of the corpora. Only corpora with naturalistic parent-child interactions were included in the final sample. To achieve this, I excluded any corpora where the investigator played an active role (e.g., interviews, defined as >10% of total utterances classified as being produced by the investigator) and school recordings (defined as >10% total utterances by teacher). Three corpora were excluded using these criteria. An additional 2 corpora were excluded for only having single-word utterances from the target child with no caregiver corresponding utterances, suggesting that they may be transcripts from an experimental task. Additionally, one more corpus was excluded for having transcripts that contained high numbers of non-English utterances.

The final set consisted of data from 67 children from 9 corpora (Forrester, Howe, Korman, Lara, Manchester, Nuffield, Thomas, Tommerdahl, Wells). This included recordings of everyday interactions (N = 2) and free-play sessions (N = 7). Each child's transcript provided between 103 and 162518 utterances (including both child-directed and child-produced utterances), totalling 979251 utterances. In the compiled dataset, 50.9% of total utterances were by the target child's mother, 39.4% were by the target child, and other individuals (father, investigator, grandparents and siblings) made up the remaining 9.7%. Common English contractions (e.g "it's") were expanded to their full form (i.e. "it is"). All words were then lemmatised using the `textstem` package (Rinker, 2018). Additionally, if a given word had abbreviations (e.g. airplane / plane) or synonyms (e.g. bunny / rabbit) that are accepted as tokens of the same item in the CDI, the frequency and co-occurrence statistics of these word tokens were summed.

CHILDES word frequency

Word frequency was obtained from the CHILDES data above by counting the number of occurrences of a given lemma, log-transformed into zipf values (Zipf, 1949). To check if child-produced utterances differed greatly from child-directed utterances, I separated CHILDES transcripts into utterances produced by the target child and utterances produced by other speakers. Frequency scores extracted from these two sources had a correlation of $.94$. As the correlation was very high, I decided not to analyse these sources separately.

CHILDES co-occurrence degree (contextual diversity)

Co-occurrence degree was also extracted from the CHILDES data. Hills et al. (2010) found that a relatively small sliding window of 5 words was the best predictor for AoA when building a measure of contextual diversity from word co-occurrence, reflecting the limitations of working memory in the developing brain. I therefore decided to use a 5-word window in this British extension. *Contextual diversity* for a given word was computed by summing the total number of unique words that the word co-occurred with in the dataset within a 5-word window.

6.2.2 Results

I investigated the contribution of the three factors (frequency, contextual diversity, associative in-degree) in predicting AoA of individual words. Frequency was transformed to zipf values, and both contextual diversity and associative in-degree were log-transformed. All mentions below of the predictors will refer to these transformed variables. To avoid the effects of influential outliers, words that were more than 3 SD from the mean of their word class in any of the three predictors were excluded from the analysis. This removed 4 nouns, 1 verb and 2 function words.

I ran linear regressions with AoA as the dependent variable using the `lme4` package (Bates et al., 2015) in R (R Core Team, 2013) for all words together and also for each word class. Notably, frequency and diversity were highly correlated ($.99$, Pearson's *r*). In all cases, frequency was a better predictor than diversity, so diversity was dropped from the models to avoid collinearity. Model fit is visualised in Figure 6.1 for comprehension and Figure 6.2 for production using the `ggiraphExtra` R package (Moon, 2021). R^2 for the models with frequency and associative in-degree are listed in Table 6.1 for comprehension. Frequency was a significant predictor for the AoA of the analysed words, but this seemed to be wholly driven by the high proportion of nouns among the sampled words. Frequency was a significant predictor for nouns but not any of the other word classes. Associative in-degree also significantly predicted AoA in the analysed words, but again there were differences between noun classes. The AoA of nouns and function words, but not verbs and adjectives, were predicted by associative in-degree. Importantly, associative in-degree explained significant additional variance for AoA of nouns and function words even after frequency was accounted for.

Models for production are listed in Table 6.2. The pattern of result for AoA in production was very similar to the pattern found for comprehension. Frequency significantly predicted AoA of nouns but not the other word classes. Associative in-degree significantly predicted AoA of nouns and function words, and explained additional variance over a model with only frequency for both of these word classes. This difference between word classes with be discussed further in the discussion section.

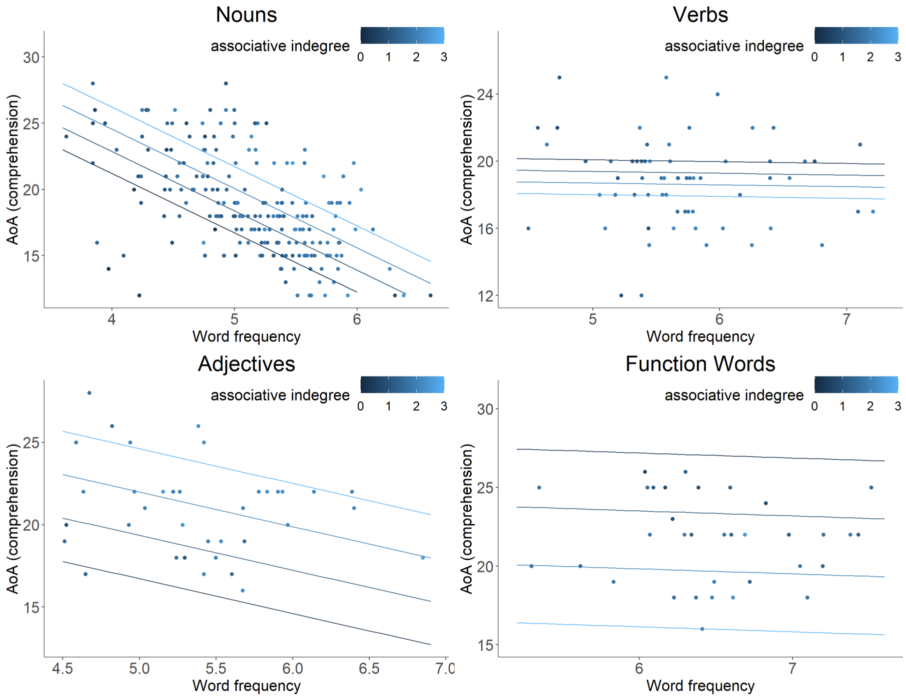

Figure 6.1: Scatterplot showing model fit, with AoA in comprehension (months) on y-axis, word frequency (zipf) on x-axis and associative in-degree (log) as colour, split by word class.

Table 6.1: R^2 of models predicting AoA (comprehension) for each word class, for frequency alone (Freq), associative in-degree alone (AI) and the increase in R^2 from adding AI as predictor after accounting for Freq.

	df	Freq	AI	ΔR^2 AI
All	341	.027**	.010*	.0008
Nouns	205	.293***	.018*	.027**
Verbs	62	$-.014$	$-.006$	
Adjectives	34	.022	.0008	
Function	32	$-.025$	.300***	.304***

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

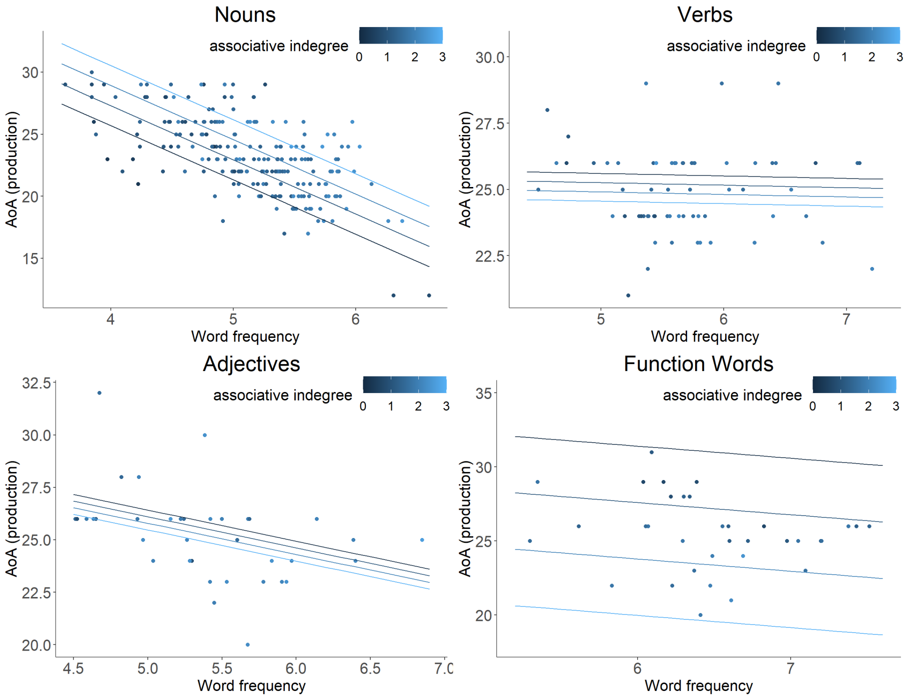

Figure 6.2: Scatterplot showing model fit, with AoA in production (months) on y-axis, word frequency (zipf) on x-axis and associative in-degree (log) as colour, split by word class.

Table 6.2: R^2 of models predicting AoA (production) for each word class, for frequency alone (Freq), associative in-degree alone (AI) and the increase in R^2 from adding AI as predictor after accounting for Freq.

	df	Freq	AI	ΔR^2 AI
All	341	.043***	.025**	.008
Nouns	205	.404***	.027**	.039***
Verbs	62	−.013	−.008	
Adjectives	34	.152*	.039	
Function	32	.010	.410***	.415***

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

6.2.3 Study 1 Discussion

This study contributes to the literature by replicating the procedures used by Hills et al. (2010) (Study 1) with an independent sample of British English toddlers and a different dataset of adult associations. I replicated their findings of significant but small R^2 for frequency and associative in-degree in predicting AoA when looking across all words together. Stronger R^2 were found within word classes, together with notable variability depending on the word class and predictor, suggesting that different word classes rely on variable mechanisms for acquisition.

Additionally, unlike the CDI used by Hills et al. which only collects production scores, the Oxford CDI includes both comprehension and production scores. This allowed me to investigate whether frequency and associative in-degree predict comprehension and production differently using data from the same children, controlling for potential confounds of developmental change. Aside from the relationship between frequency and AoA for adjectives that reached significance for production but not comprehension, the pattern of results in this study was very similar between comprehension and production.

Frequency and AoA

When looking at each word class separately, nouns were the only word class where both predictors were significantly correlated to AoA. However, when both predictors were entered in a regression together, frequency accounted for a dominant amount of variance by itself. This supports the extensive literature which argues that repeated exposure to words (and by proxy word-object mappings) predicts more successful word learning. The more times a word-object pair occurs, the easier it is to learn. The especially strong effect of frequency on the AoA of nouns relative to other word classes is a common finding in the literature (Goodman et al., 2008; Hills et al., 2010; Braginsky et al., 2019).

Unexpectedly, frequency was very bad at predicting acquisition order for verbs. This poor predictive power of frequency is surprising considering findings in the literature of frequency predicting verb acquisition (L. R. Naigles & Hoff-Ginsberg, 1998). It also contrasts with Hills et al. (2010)'s findings of significant predictors. It is possible that the profile of the dataset may have masked any effect of frequency. The 64 verbs included in the analyses had a narrower spread of AoA in comprehension (mean = 18.78, IQR = 3) as compared to the 207 nouns (mean = 18.79, IQR = 5.5). Further exploration with datasets including a wider distribution of verbs is required to identify whether this poor effect of frequency is meaningful or simply a limitation of the data.

The literature is more mixed for function words. Infants have been shown to be sensitive to frequency as a cue to differentiate function words from other word classes (Hochmann et al., 2010). However, when investigating AoA within the word class, Braginsky et al. (2019) found that children's knowledge of function words was predicted better by word length and sentence complexity. This was in contrast to nouns and predicates which were strongly related to frequency. Braginsky et al. proposed that low sentence complexity could allow learners to decode a function word's meaning more easily. Hidaka (2013) proposed a computational model for the acquisition of function words that is insensitive to frequency, instead suggesting that function words are learnt through cognitive processes related to inference. This is again contrasted with the acquisition patterns of nouns, verbs and adjectives, which Hidaka found were best predicted by a model of cumulative learning. The divide between function words and other word classes presented in the above-mentioned studies is consistent with the findings reported in this chapter.

Collinearity of frequency and contextual diversity

A strong correlation between frequency and contextual diversity as derived from CDS was found both across all words and within word classes. Shaoul and Westbury (2006) addressed the issue that very high frequency words tend to have higher scores in word co-occurrence measures simply through chance co-occurrences, in the context of the hyperspace analog to language (HAL) model. This issue is particularly relevant for corpus studies of infant language, where the earliest learnt words are often highly frequent. However, applying a simple standardisation procedure as in Hills (2013) where the contextual diversity score for a word was divided by the word's frequency did little to reduce this collinearity in my data. The resulting correlation between frequency and the standardised score (both log-transformed) was still very high at $-.985$. More exploration is needed to identify a suitable standardisation method to quantify contextual diversity without the influence of chance co-occurrences in higher frequency words. This will allow us to study whether contextual diversity plays an independent role in facilitating word learning outside of frequency effects.

Additionally, considering Chang and Deák (2020)'s findings that both adjacent and non-adjacent co-occurrences contribute unique variance to predicting AoA, a weighted co-occurrence system as applied by HAL and Shaoul and Westbury's (2006) HiDEx models may allow us to explore the effect of contextual diversity in CDS at a more fine-grained level than the binary co-occurrence matrix used by Hills et al. (2010) and also the present study.

Semantic associations and AoA

Both nouns and function words showed a significant relationship between associative in-degree and AoA, with the predictor explaining significant variance even after accounting for frequency. This effect was small for nouns, while associative in-degree was by far the best predictor for function words out of the available options.

Like Hills et al. (2010), I found a small correlation between associative in-degree and contextual diversity as derived from CHILDES when looking at all words together ($.25$, Pearson's *r*). As suggested by Hills et al., diverse semantic contexts could make a word more salient in the learning environment and support the disambiguation of meaning. There were stronger correlations within word classes (Nouns $.56$, Verbs $.30$, Adjectives $.48$) with the striking exception of function words ($-.003$). This suggests that while contextual diversity may explain part of the relationship between associative in-degree and AoA of nouns, there is something different at play with function words.

I suggested in the introduction of this chapter that associative in-degree rep-

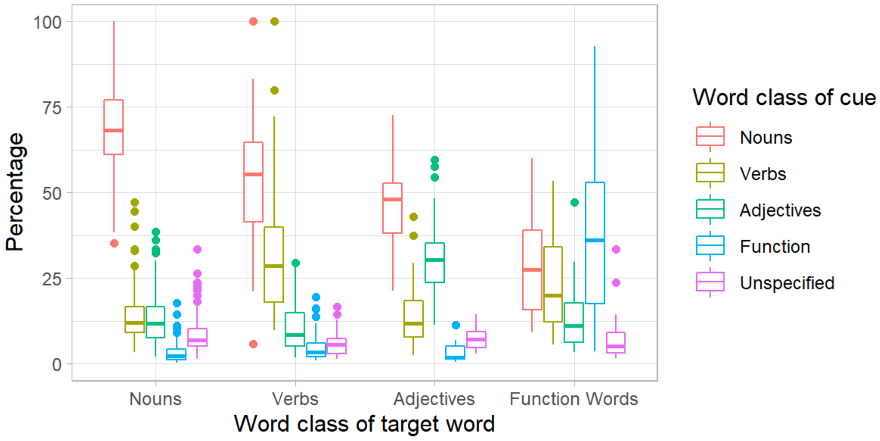

Figure 6.3: Boxplots showing the distributions of the word class of the cue words that elicit a given response word in the Small World of Words database, with the response words also split by word class.

resented the richness of a word's semantic environment with a focus on paradigmatic relationships. An initial exploration of the composition of associations by word class supports this theory. On average, 37.1% of cues that elicited the sampled function words were other function words (28.9% cues were nouns). In comparison, for the sampled nouns, on average 69.5% of the cues were other nouns (only 3.4% were function words). The distribution of word classes for words that have associative links in the Small World of Words database is shown in Figure 6.3. Most frequent cue-response types include antonyms like "you"–"me", "this"–"that"; synonyms like "beneath"–"under", "additional"–"more"; and phrases like "thank"–"you", "upside"–"down". Unlike concrete nouns which have obvious referents, the learning of function words is dependent on the interpretation of its meaning from the context it appears in. Rich semantic associations may facilitate easier inference of a function word's meaning. Steyvers and Tenenbaum (2005) proposed that words that are connected to many different semantic neighbourhoods are more semantically distinct than words that are less connected, making their meanings easier to interpret. Further work needs to be done to study the underlying mechanisms represented by the associative in-degree measure, and whether paradigmatic and syntagmatic relationships between words contribute differently to word acquisition and inference of meaning.

6.3 Study 2: Associative vocabulary structure of toddlers

Given the findings that nouns and function words with higher associative in-degree have earlier AoA than those with lower associative in-degree, how does this affect the structure of early vocabulary? To study this question, I quantified the structural properties of the words known by a given child by building semantic networks. Analysis of the structural properties of semantic networks can provide us with insight on the degree of connectivity held between different sets of words. If toddlers' word learning follows a pattern where words that are highly connected to other words are learnt earlier, I would expect to see higher connectivity between early words relative to a network of equal size where nodes are randomly selected. To evaluate a semantic network as a whole, we can average the degree or path lengths across all nodes in the network, and also calculate the clustering coefficient of the nodes (i.e. how many nodes are fully connected in groups of three). The average degree of words in the network tells us about the amount of connectivity the words have with each other – if the words are semantically-unrelated (e.g. "bread", "jump", "blue"), each word would have low degree and result in low average degree. Conversely, if words were closely related (e.g. "bread" is connected to “oven” and “bake” and “brown”), words would display high degree from the many connections the words int the network have with each other. A similar pattern is shown by average path length, where semantically-unrelated words result in long average path lengths (e.g. as “bread” and “blue” have no direct semantic relationship, the two words would need to rely on an intervening path of “bread”–“brown”–“blue”). Finally, clusters in the network tell us about the closeness of contexts that words occur in. If the words that a node word is connected to are also connected to each other, this would suggest that the words all occur across similar contexts (e.g. “bread”, “cake”, “biscuit” are all linked). If there are few connections between words that a node word is connected to, this suggests that the node word occurs in various different contexts that rarely or never overlap (e.g. “brown” is linked to “bread” and “wood”, but “bread” and “wood” are not related to each other). Semantic networks for language have been found to have small-world structure, i.e. shorter average path length between nodes and more highly clustered neighbourhoods relative to random networks (Steyvers & Tenenbaum, 2005). Steyvers and Tenenbaum consistently found this structure across semantic networks built using 3 different data sets (free associations; Roget’s thesaurus for semantic categories; and WordNet for semantic relations). This suggests that there is a systematic pattern linking the words that are frequently used in a language. The early vocabulary of children have also been shown to exhibit small-world structure (Hills et al., 2009; Hashemikamangar, Bakouie, & Gharibzadeh, 2020). Small-world structure in early vocabulary suggests that toddlers’ learning of words follows a systematic pattern where they prefer to learn groups of related words (e.g. “bread”, “milk” and “eat”), as opposed to unrelated words (“bread”, “jump” and “green”). This pattern of learning is logical when considered in relation to developmental milestones and the physical learning environment. Young infants with low mobility are exposed to limited words and experiences ("mummy", "milk", "hug") which later expands as their experiences broaden (new food items like "carrot", more actions like "jump", new places like "school", etc). Child-directed speech has been found to be more repetitive and associative than adult-directed speech (Hills, 2013), supporting the idea that toddlers' learning environment contains these small-world properties.

Measures of small-world structure in vocabulary have been used to compare the vocabulary compositions of different children to identify possible sources of divergence. Low connectivity observed in an early learners' vocabulary may be indicative of a broader problem. Beckage et al. (2011) compared monolingual late-talkers to their typically-developing peers. They found that monolingual toddlers classified as late talkers (in the lowest 20th percentile of vocabulary size for their age group) show lower semantic connectivity in their vocabularies (lower clustering, lower in-degree and longer distance between nodes) than typically-developing toddlers matched for vocabulary size.

Bilson et al. (2015) compared the semantic networks made from vocabularies of monolingual and bilingual toddlers. The preferential acquisition model (Hills et al., 2009) fitted best to explain the vocabulary growth of both monolingual and bilingual toddlers, suggesting that both groups share the preference of learning words that are highly connected to other words in the learning environment. However, comparing measures of small-world structure between the two groups presented a more complicated pattern. Similar to late talkers, bilinguals' networks showed longer average path lengths than the networks of same-age monolinguals.

In contrast, bilinguals' average in-degree was higher than that of monolinguals, suggesting that bilinguals may be even more sensitive than monolinguals to the highly-connected words proposed to be learnt first under the preferential acquisition model. Bilson et al. (2015) suggested that this stronger preference for high in-degree words in bilinguals may be the result of facilitation from words known in their other language.

The aims of Study 2 were three-fold – I first compared the semantic networks built from monolingual toddlers' vocabulary against random networks to probe if the small-world properties observed by Steyvers and Tenenbaum (2005) and Hills et al. (2009) can also be observed in an independent sample of UK monolingual toddlers. I then compared the structural properties of toddlers whose vocabulary size falls in different percentiles relative to their same-age peers, in an attempt to extend the findings of Beckage et al. (2011) that there are differences in structural properties between the vocabularies of typically-developing toddlers and late talkers. Finally, I compared the semantic networks of a sample of bilingual toddlers against monolingual toddlers of the same age. Differences between monolinguals and bilinguals' vocabulary networks have previously been found by Bilson et al. (2015). Studying the semantic network structure of bilingual toddlers' vocabulary can shed more light on cross-linguistic influence on vocabulary growth.

6.3.1 Methods

Participants

I constructed associative networks for each of the toddlers in the sample from Study 1 described in Section 6.2.1 of this chapter. Edges in the network were derived from the free association data from the SWOW database also described in Section 6.2.1. An edge is defined by two words that have been associated as a cue and response pair by participants in the SWOW free association task. The network was defined as a directed network, with edges directed from cue words towards response words. Networks were generated independently for each toddler using the `igraph` package (Csardi & Nepusz, 2006) in R. Each node in the network represented a word that the toddler was reported to understand in the Oxford CDI. As with Study 1, I excluded words about time, sounds, games and routines, compound words, and words that occur in multiple Oxford CDI categories. This left 348 words.

Toddlers' vocabulary sizes in comprehension from these 348 words ranged from 3 to 348 (mean 201.6). To calculate percentiles of vocabulary size, I grouped toddlers into age groups by months. I then ranked toddlers' vocabulary sizes into percentiles compared to toddlers in the same age group. I finally grouped toddlers into 5 groups by their vocabulary size percentile, corresponding to 0-20th percentile, 20-40th percentile, 40-60th percentile, 60-80th percentile and 80-100th percentile, respectively. To avoid floor and ceiling effects on the analyses of network structure, toddlers with vocabulary size smaller than 50 words and those with vocabulary size larger than 300 words were excluded from further analyses. This left 1403 toddlers, with mean age of 19.6 months old (range 12;0–32;0 months), and mean vocabulary size of 180.6 words. After these exclusions, 246 toddlers fell in the 0-20th percentile for their age group, 307 in the 20-40th percentile, 330 in the 40-60th percentile, 299 in the 60-80th percentile and 221 in the 80th percentile.

Finally, I compared the semantic networks of 15 to 25-month-old monolingual toddlers and bilingual toddlers, using a subset of the sample reported in Chapter 2. This sample was collected between 2020 and 2022 for both groups. To avoid floor and ceiling effects in vocabulary size affecting network structure, I restricted analyses to toddlers with English vocabulary size between 50 and 300 words, and also conceptual vocabulary size between 50 and 300 words. For monolinguals, conceptual vocabulary is equal to their English vocabulary. For bilinguals, conceptual vocabulary counts all concepts known by a bilingual toddler regardless of which language the concept is known in. The sample size, after the above filtering, was 247 monolinguals (mean age 20.2 months) and 176 bilinguals (mean age 20.1 months). I compared groups on both English vocabulary networks and conceptual vocabulary networks. I investigated whether associative patterns in English (the associative networks in this study were built using English data from the SWOW free association norms) predicted both toddlers' vocabulary acquisition in English and their overall acquisition of concepts.

Random acquisition networks and Erdős–Rényi networks

To evaluate whether the vocabulary networks of the sampled toddlers show small-world properties, I compared the semantic networks of toddlers' vocabulary against two types of random networks – random acquisition networks and Erdős–Rényi (ER) graphs (Erdős & Rényi, 1960). *Random acquisition networks* were built from words randomly sampled from the CDI. This sampling was conducted for every possible vocabulary size in the sample (range 1–365) with 300 iterations each. Network connections were specified using the associative network described above. Random acquisition networks tell us about the inherent structure present in language, but disregards any patterns that guide words' age-of-acquisition. On the other hand, *ER graphs* are networks with randomly assigned edges. As with random acquisition networks, I generated ER graphs for each possible vocabulary size, with 100 iterations. ER graphs were generated using the `erdos.renyi.game` function in the R `igraph` package (Csardi & Nepusz, 2006). I specified the number of edges for each of these ER networks to match the average number of edges of the random acquisition networks as specified above. In the following analyses, ER networks served as a baseline for connectivity within a network, as they have equal sizes to toddler networks but without language-specific structure. By comparing toddlers' vocabulary networks against these two types of random networks, I evaluate the degree to which toddlers' vocabulary growth is guided by the structure of language.

Network statistics

To describe the overall structure of the networks, I utilised three measures of semantic network connectivity commonly used in the literature – average in-degree, clustering coefficient and average geodesic distance of all nodes in the network. These network statistics were calculated for each child.

The *in-degree* of a node is calculated by the number of edges it has directed towards it. In the case of associative networks made from free-association norms, each node is a response word and subsequent in-degree is defined as the number of unique cues that elicit a given response word. In-degree has commonly been used to evaluate directed semantic networks, particularly for associative networks built from free association tasks, as out-degree (the number of different responses given for a particular cue word) can change depending on the structure of the task and sample size (Steyvers & Tenenbaum, 2005). The average in-degree of all nodes in the network is used as a measure of the network's overall connectivity. Networks with higher average in-degree are said to be better connected, as the words are directly linked to many other words in the network.

Clustering coefficient is a probabilistic measure of the connectivity of the network. Two nodes that are connected to each other by an edge are called neighbours. Clustering coefficient represents the probability that two nodes that are neighbours with a given node are also connected to each other (thereby forming a triangle of connections between the three nodes). The clustering coefficient of a single node is calculated using the number of completed triangles made between its neighbours over the total number of possible triangles that could have formed if all neighbours were connected to each other. To provide a numeric example, if a node has 3 neighbours, there are a maximum of 3 triangles that could form, and if only 1 pair is connected to each other, the clustering coefficient for that node is 1/3.

Geodesic distance is calculated using the shortest path length between two nodes in the network. For example, if Node A is not connected to Node B but both nodes are connected to Node C, the shortest path length between Node A and B is 2 (A to C to B). A network with a shorter geodesic distance (short average path lengths) is considered to be better connected than a network with longer geodesic distance, as this means that a larger number of nodes are connected directly to each other or via only one intermediary.

The network statistics for random acquisition networks of each vocabulary size were averaged across its 300 random iterations. Likewise, the network statistics for ER networks of each vocabulary size were averaged across its 100 iterations.

6.3.2 Results

Comparing to random networks

I ran linear regressions for each network statistic, comparing toddlers' vocabulary networks against random acquisition networks and ER networks. Toddlers' network was the reference level for all these models. As these statistics are sensitive to the size of the network, I included the number of words in the networks as a covariate in each regression. If toddlers' vocabulary acquisition is sensitive to the structure of the language being learnt, toddlers' early vocabulary should show small-world structure, i.e. higher in-degree, higher clustering coefficient and shorter geodesic distance as compared to both random acquisition and ER networks. Small-world structure would suggest that early-learnt concepts are semantically-related to each other.

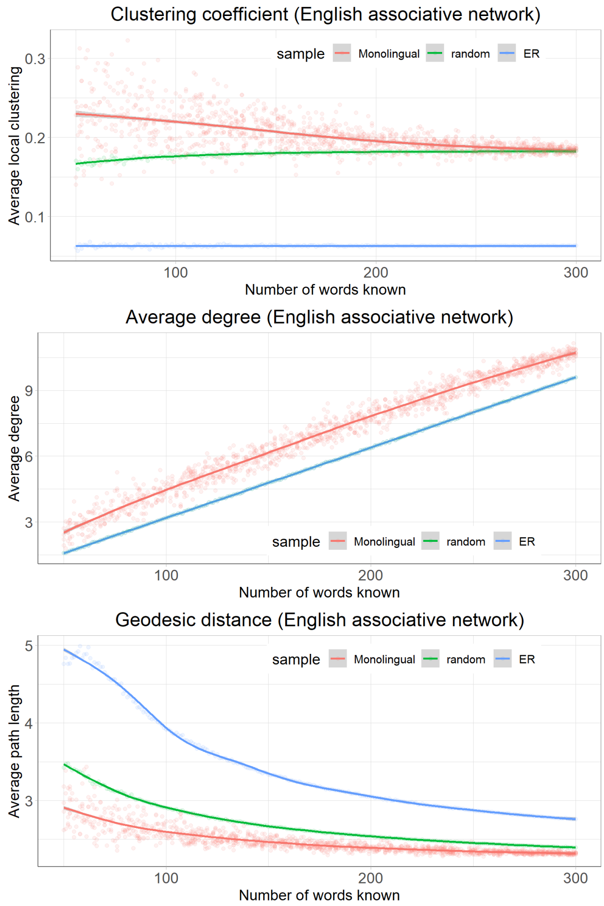

Figure 6.4: Scatterplots showing network statistics from an English associative network on the y-axis, with number of vertices on the x-axis, comparing vocabulary networks of monolingual toddlers against random acquisition networks (random) and Erdős–Rényi networks (ER) of the same size.

As predicted, associative networks built using toddlers' vocabulary displayed small-world properties relative to both ER networks and random acquisition networks (Figure 6.4). Toddlers' vocabulary networks showed higher clustering coefficient than both ER networks (t = -109.56, $p < .001$) and random acquisition networks (t = -19.13, $p < .001$). Toddlers' vocabulary networks also showed higher average in-degree than both ER networks (t = -59.4, $p < .001$) and random acquisition networks (t = -59.4, $p < .001$). Finally, toddlers' vocabulary networks showed shorter geodesic distance than both ER networks (t = 68.25, $p < .001$) and random acquisition networks (t = 15.29, $p < .001$). This supports the idea that toddlers are sensitive to the associative structure in the language environment and preferentially learn words with high connectivity.

ER networks are characterised by having a specified number of nodes and number of edges, but with the edges randomly arranged without a systematic pattern. As a result, the clustering coefficient is typically very low. Therefore, it is unsurprising that the clustering coefficient for ER graphs is much lower than both toddler networks and random acquisition networks, with a mean clustering coefficient of 0.0663 and a very small standard deviation of 0.000850. The small standard deviation of ER network's clustering coefficient can be linked to the network's ratio of edges to nodes. In the present study, ER networks were built for each vocabulary size by using the average number of edges of random acquisition networks of the corresponding vocabulary size. This is the reason for there being no difference between ER networks and random acquisition networks for average degree. The average number of edges from random acquisition networks show linear increase (slope of 11.1) relative to the number of nodes, with a very high correlation of .969 between the number of nodes and number of edges. The proportional increase of edges with the increasing number of nodes resulted in a stable average clustering coefficient across different network sizes.

Another pattern that can be observed in Figure 6.4 is that clustering coefficient for toddler networks starts out higher than random acquisition networks at small vocabulary sizes then appears to converge with random acquisition networks at larger vocabulary sizes. Similarly, geodesic distance of toddler networks is smaller than random acquisition networks at low vocabulary sizes but similar at larger vocabulary sizes. However, when attempting to interpret this pattern, it is important to be aware of some limitations in the methodology. The vocabulary used to build these networks is obtained from the CDI, which is a finite word list. This finite word list has effects on the probability of random sampling. A random acquisition network that randomly samples only 50 words out of 348 has low chance of overlap with the words known by a toddler who has 50 words in their vocabulary. However, a random acquisition network that samples 300 out of 348 possible words has a high chance of words overlapping with the words known by a toddler with a vocabulary size of 300 words. Therefore, the larger the vocabulary size that is analysed, the set of words in toddler networks and random acquisition networks become more likely to be identical. As such, the observed convergence of network statistics between toddler networks and random acquisition networks should not necessarily be considered as a characteristic of toddlers' vocabulary growth. Due to this confound, I have removed vocabulary sizes between 301 and 348 words from the analysis, but the pattern of convergence can be observed throughout the sampled range of vocabulary sizes. It is also possible that toddlers, especially those with high vocabulary sizes, also have many more words in their vocabularies that are not listed in the CDI. As any additional words known by the toddlers are also likely to be densely connected to the words in the CDI, the inclusion of these words is likely to change the network statistics of toddler networks. Further research with a larger word list or improved sampling methodology can help us tease apart the patterns that reflect toddlers' vocabulary development and those that may be caused by limitations in methodology.

Comparing across percentiles

I then compared the network structure of toddlers across different vocabulary size percentile groups. As before, I ran linear regressions for each network statistic, using the number of words as a covariate. The group with the lowest percentile (0 to 20th percentile) was the reference group. Figure 6.5 visualises the relationship between the number of words and the three network statistics for each group.

The vocabulary networks of toddlers in the 0-20th percentile did not show significant difference in clustering coefficient from toddlers in the 20-40th percentile (t = 0.525, p = .600) and 40-60th (t = 0.957, p = .339). However, the 0-20th percentile group showed clustering coefficients that were marginally lower than 60-80th percentile group (t = 1.705, p = .0884) and significantly lower than 80-100th percentile group (t = 2.536, p = .0113).

Toddlers in the 0-20th percentile did not show significant difference in average in-degree from the 20-40th group (t = 0.683, p = .495), 40-60th group (t = -0.768, p = .443), 60-80th group (t = -0.471, p = .638) and 80-100th group (t = -0.764, p = .445). They also did not show significant difference in geodesic distance from the

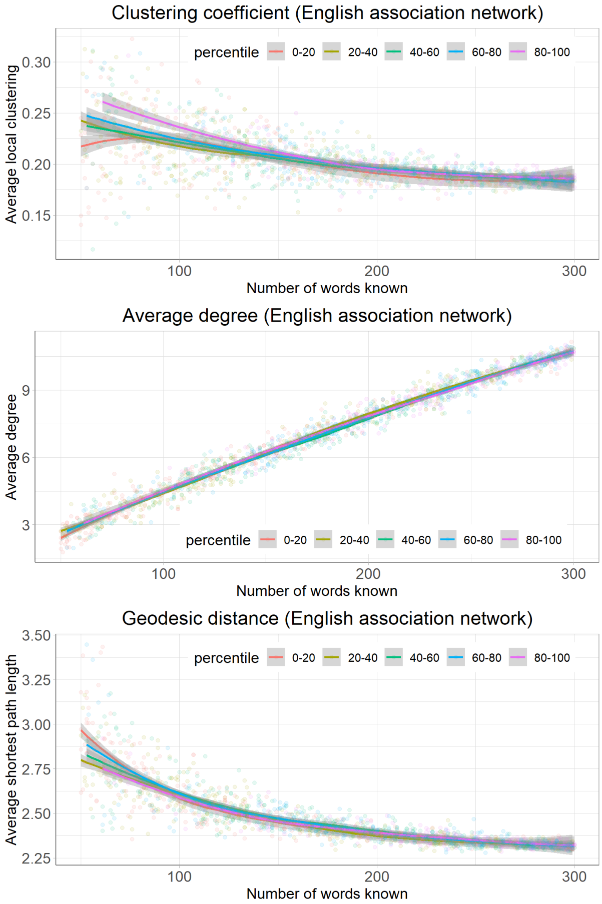

Figure 6.5: Scatterplots showing network statistics from an English associative network on the y-axis, with number of vertices on the x-axis, comparing vocabulary structure across monolingual toddlers whose vocabulary size falls in different percentiles for their age.

20-40th group ($t =$ -1.772, $p =$ .0766), 40-60th group ($t =$ -0.321, $p =$ .748), 60-80th group ($t =$ 0.043, $p =$ .966) and 80-100th group ($t =$ 0.661, $p =$ .509).

Taken together, there was a difference between the semantic networks of toddlers whose vocabulary sizes fall in the lowest percentiles for their age group and toddlers whose vocabulary sizes fall in the highest percentiles, suggesting that toddlers whose vocabulary development lags behind their peers may be less adept at utilising the associative structure of language to guide their learning. However, this difference was not consistent across all three measures of small-world structure, only being significant for clustering coefficient. This suggests that differences between groups are small. Importantly, even toddlers in the lowest vocabulary size percentiles show small-world structure in their vocabulary relative to random networks.

Comparing across bilinguals and monolinguals

The final analysis that I conducted in this chapter compared the network structure of bilingual toddlers and monolingual toddlers' vocabulary. Again, I ran linear regressions for each network statistic, using the number of words as a covariate. The monolingual group was the reference group. Figure 6.6 visualises the relationship between toddlers' English vocabulary size and the three network statistics for monolingual and bilingual groups. The English vocabulary networks of bilingual toddlers showed higher average degree (t = 4.506, p < .001) and lower geodesic distance (t = -1.974, p = .0491) than the English vocabulary of monolingual toddlers. Clustering coefficient was not significantly different between groups (t = 0.65, p = .516). These statistics indicate that bilingual toddlers' vocabularies display stronger small-world properties than that of monolinguals. This pattern is interesting as it suggests that bilinguals may be relying even more on associative patterns to guide their word learning than monolinguals.

Figure 6.7 visualises the relationship between conceptual vocabulary size and the three network statistics for monolingual and bilingual groups. Similar to En-

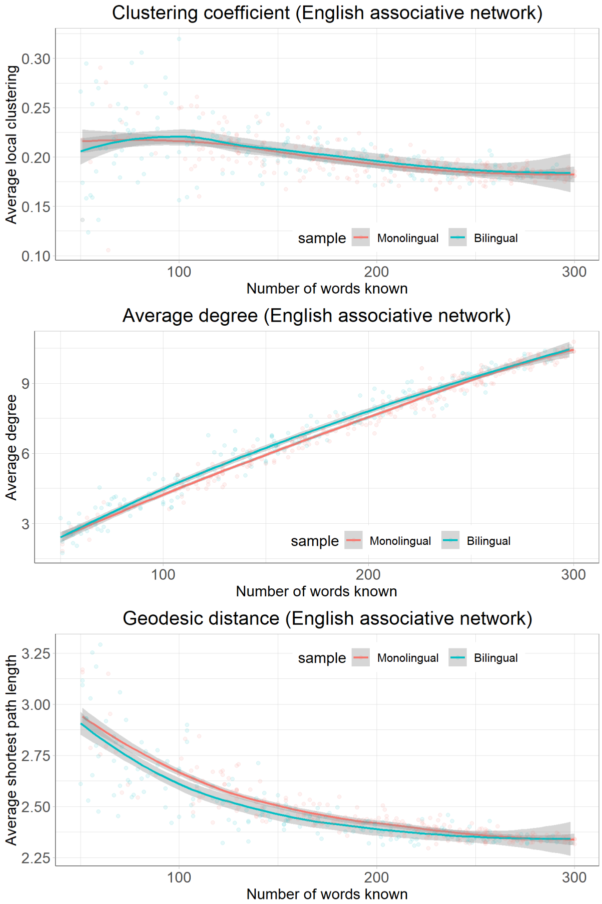

Figure 6.6: Scatterplots showing network statistics from an English associative network on the y-axis, with number of vertices on the x-axis, comparing vocabulary structure between bilingual and monolingual toddlers (English vocabulary).

glish vocabulary networks, bilinguals' vocabulary networks showed stronger small-world properties than that of monolinguals. For conceptual vocabulary, the difference was significant for all three metrics. The conceptual vocabulary networks of bilingual toddlers showed higher clustering coefficient ($t = 2.203$, $p = .0282$), higher average degree ($t = 2.99$, $p = .00295$) and lower geodesic distance ($t = -2.466$, $p = .0141$) than the vocabulary networks of monolingual toddlers.

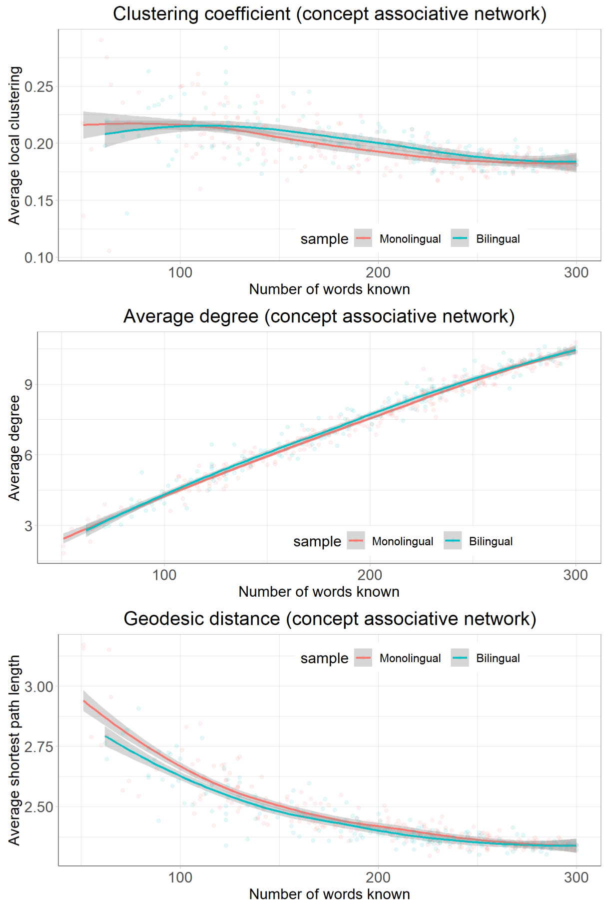

Figure 6.7: Scatterplots showing network statistics from an English associative network on the y-axis, with number of vertices on the x-axis, comparing vocabulary structure between bilingual and monolingual toddlers (conceptual vocabulary).

The consistent difference between groups across both English vocabulary and conceptual vocabulary suggests that bilinguals' vocabulary acquisition may be influenced by cross-linguistic influences, making them differ from the pattern observed in monolinguals.

6.3.3 Study 2 Discussion

The patterns observed in Study 2 support the existing literature claiming that toddlers are sensitive to the associative structure of language and that these features affect their vocabulary growth. Networks built using toddlers' vocabularies exhibit small-world structure, similar to the adult lexicon as characterised by Steyvers and Tenenbaum (2005).

Structural properties of the lexicon have been linked to behavioural responses in language tasks. The centrality of a node within the semantic network has been found to be negatively correlated to adult participants' reaction time in lexical decision and naming tasks, with faster reaction times for words that are more central (defined as having a high in-degree). Semantic connectivity can also affect judgements of semantic similarity, and subsequently classification of concepts into associated categories. The path length between two nodes in a semantic network (i.e. how many other nodes intervene between the two given nodes) predicts response time in judgements of semantic relationship between word pairs, and also predicts accuracy in free recall and cued recall (Kenett, Levi, Anaki, & Faust, 2017). Word pairs with a close semantic relationship (defined computationally as having short path lengths) are easier to retrieve from memory.

Comparing across percentiles

I also show a difference in clustering coefficient between toddlers whose vocabulary size falls in the lowest percentiles for their age (0-20th) and those at the highest percentiles (80-100th), even after the number of words known was controlled for. This suggests that the semantic networks of toddlers whose vocabulary sizes are much smaller than the average for their age group have less densely-populated connections between known words than the networks of toddlers whose vocabulary are in the high range for their age group.

Despite the difference observed for clustering coefficient, there was no significant difference between groups for both average in-degree and geodesic distance. Importantly, even the 0-20th percentile group in the sample showed a significantly higher clustering coefficient, higher average in-degree and lower geodesic distance than both random acquisition networks and ER networks. This suggests that all toddlers in the sample, regardless of percentile, were able to make use of the structure of language to guide their vocabulary acquisition. However, these results notably differ from the pattern observed by Beckage et al. (2011), who found significant differences between typically-developing toddlers and late talkers for all three statistics. The late talkers in Beckage et al.'s sample additionally had a lower clustering coefficient compared to random acquisition networks. One possibility for the difference between the two samples is that the sample reported in this chapter excluded toddlers with diagnosed language or learning difficulties. The percentiles in the present study may therefore only represent percentiles within the population of typically-developing toddlers. The lack of observed group differences for in-degree and geodesic distance may therefore be due to a smaller variance in vocabulary size than that of Beckage et al.'s sample.

It is important to note that all analyses conducted in Study 2 controlled for the number of words in the network. Including this covariate is crucial because the number of nodes in a network has strong effects on the nodes' degree and path lengths between nodes (Erdős & Rényi, 1960), which also influences measures of clustering coefficient. However, by using the number of words as a covariate in the linear regression, I are inevitably comparing older toddlers with comparatively small vocabulary size for their age against younger toddlers with comparatively large vocabulary size for their age. This has implications for the interpretations of any group differences (or lack thereof). As children grow older, their environment and experiences tend to change as well. A 12-month-old is unlikely to know what "scissors" are, given that they typically wouldn't be allowed near the potentially-dangerous item. On the other hand, an older toddler may be allowed to use child-safe plastic scissors, broadening both their knowledge of the world and their word knowledge. The complexity of child-directed speech from parents has also been found to increase together with their child's age and speech complexity (Kunert, Fernández, & Zuidema, 2011). There may therefore be qualitative differences in both the physical and language environment experienced by toddlers of different ages, which can in turn affect which words their vocabulary is made up of. Nevertheless, the lack of observed difference between percentile groups in the present study for average in-degree and geodesic distance suggests that toddlers are still preferentially learning words with similar properties in the language environment (e.g., words with high in-degree), even for the toddlers who show slower vocabulary growth.

Comparing across bilinguals and monolinguals

Moving to the analyses comparing monolinguals and bilinguals, I found that bilinguals showed higher in-degree and shorter geodesic distance in their English vocabulary networks than monolinguals with the same vocabulary size. Clustering coefficient was not significantly different between groups. When comparing networks built from conceptual vocabulary, a similar pattern of higher in-degree and shorter geodesic distance was again found for bilinguals relative to monolinguals. Bilinguals' conceptual vocabulary also showed significantly higher clustering coefficient than monolinguals.

This pattern of results is partially consistent with previous findings from Bilson et al. (2015). Bilson et al. found higher average in-degree for their sample of bilingual toddlers relative to monolingual toddlers when analysing toddlers' English vocabulary networks. They also no significant difference in clustering coefficient, which was also the case for English vocabulary networks in the present study. However, in contrast to the shorter path lengths found in the present study, Bilson et al. found longer path lengths for bilinguals than monolinguals. It is difficult to be sure if the different results found for average path length between the present study and Bilson et al. (2015) is a outcome of the sample or the methodology. While both studies used free association data, the present study used the Small World of Words database (De Deyne et al., 2019) while Bilson et al. used the University of South Florida Free Association Norms (Nelson et al., 2004). The present study also used the Oxford CDI (Hamilton et al., 2000) to measure vocabulary in comprehension and production (however, only comprehension has been analysed due to time constraints), while Bilson et al. used the MacArthur–Bates Communicative Development Inventory (MCDI) Toddler Form (Fenson et al., 1994) to measure vocabulary in production only. Future studies on semantic networks using Oxford CDI data can extend the analyses reported in this chapter to the available production data, for direct comparability to the data from the MCDI Toddler Form. Further attempts to identify points of overlap or divergence between these methodologies, along with attempts to replicate comparisons between bilinguals and monolinguals, can help us better understand patterns in bilinguals' vocabulary growth. Despite group differences, the observation that both monolinguals and bilinguals' vocabulary display small-world properties across both the present study and Bilson et al.'s (2015) study indicates that both groups follow a similar learning trajectory where highly-connected words are learnt earlier.

6.4 General Discussion

Study 1 in this chapter extends prior research on frequency, contextual diversity and associative diversity effects on monolingual AoA of early vocabulary. Frequency was the best predictor for nouns, with associative in-degree explaining a small but significant amount of unique variance. Meanwhile, associative in-degree was the strongest predictor for acquisition order of function words. The strong relationship between word frequency and nouns is consistent with previous findings in the literature, as is the difference in predictors between function words and other word classes. However there is still much work to be done to uncover how these predictors directly impact learning mechanisms in infant language learners, and why they are found to have variable strength for predicting acquisition of different word classes. In Study 2, I went further to quantify the structural properties of early vocabulary as revealed using analyses of associative networks. While the analysis in Study 1 informs us about the features of early-learnt words in relation to the learning environment, Study 2 tells us about how these early-learnt words are connected to each other. Findings in Study 2 that toddlers' vocabulary networks show small-world properties relative to random networks supports the idea that toddlers can draw on associative patterns between words in the language to support their vocabulary learning. I also compared the structural properties of the networks built from the vocabularies of toddlers whose vocabulary sizes fall in different percentiles. Significant differences were found between toddlers in the lowest percentiles for their age group (0–20th) and toddlers in the highest percentiles (80–100th), but only in clustering coefficient. This observed difference in clustering coefficient suggests that toddlers whose vocabulary size lags behind their peers may learn words less systematically, instead of acquiring words in associative clusters. Nevertheless, this difference in associative vocabulary structure was only found in one of the three small-world measures, indicating that differences between groups was very small. Additionally, toddlers across all percentile groups showed small-world properties relative to random networks, suggesting that regardless of whether toddlers had high or low vocabulary size, the vocabulary development of toddlers in this sample followed the expected pattern of high connectivity between words.

Finally, in the last part of Study 2, I explored the effects of bilingual language experience on vocabulary development, by comparing the associative structure of vocabulary known by bilingual toddlers and monolingual toddlers. The vocabulary of bilingual toddlers showed higher clustering coefficient, higher in-degree and shorter geodesic distance compared to that of monolingual toddlers. This pattern was consistent across both English vocabulary and conceptual vocabulary. This suggests that bilinguals may rely more on associative patterns than monolinguals. These results are partially consistent with the findings from Bilson et al. (2015), who likewise found patterns suggesting that bilinguals' vocabulary acquisition was more strongly influenced by associative statistics than monolinguals. This is interesting when considered in relation to bilingual word learning. Common concepts across languages (i.e. translation equivalents) means that associative properties of languages should have some similarities. At the same time, associations can still differ between languages depending on cultural norms and language use. For example, while an association between "finger" and "hand" should be consistent across most languages, an association between "duck" and "bathtub" would be restricted to cultures where a rubber duck is a common bath-time toy. Given these potential differences between languages, further research investigating bilingual network structure both with same-language edges and cross-linguistic edges may tell us more about the acquisition of translation equivalents. A preference towards learning translation equivalents, as previously discussed in Chapter 2 may also explain some of the differences between the vocabulary structure of monolingual and bilingual. This possibility was also raised by Bilson et al. (2015). Additionally, for bilinguals, there may be context-dependent language choice, where a certain category of words may be heard mostly in one language (e.g. bath-related words in the AL at home), while other categories are heard in the other language or both (e.g. animal words from book reading in both languages). The language environment experienced by bilinguals is complex – future research is needed to disentangle the features that result in the strong associative small-world properties observed in bilingual toddlers' vocabulary.

In this chapter, I showed that the vocabulary development of both monolingual and bilingual toddlers follows a trajectory that can be linked to the associative patterns present in language. Additionally, cross-linguistic language experience not only exerts influence on phonological connections in toddlers' vocabulary (as discussed in previous chapters) but also semantic connections. These findings reveal another characteristic of bilinguals' vocabulary, but the mechanism underlying these patterns are still under-explored and warrant further research.

Chapter 7

Conclusion

This thesis covered topics of vocabulary development and lexical organisation in bilingual and monolingual toddlers. I investigated the effect of phonological connections between words in the lexicon on bilingual word learning, focusing on cross-linguistic phonological similarity. I also studied semantic connections between words and the learning environment.

The primary contribution of this thesis was studying the vocabulary development trajectories of bilingual toddlers growing up in the UK, which has a dominant majority language of English. In this population of bilinguals, most of their exposure to their minority language would be received in the home from parents and through less frequent interactions with extended family members during visits or video calls. The language environment faced by bilingual toddlers in this population is both quantitatively and qualitatively different from bilinguals growing up in more multilingual communities, for example in French-speaking regions of

Canada where many people speak both English and French. An English-French bilingual growing up in the UK is likely to have limited exposure to French outside of the home, and may receive less exposure to French overall as well. Studying the phonological and semantic connectivity of bilingual toddlers' vocabulary for this population of bilinguals can tell us more about the development of the lexicon under circumstances where the language exposure may be imbalanced and context-dependent.

Several aspects of UK bilingual toddlers' vocabulary growth were evaluated in this thesis. Firstly, the vocabulary sizes of bilingual toddlers were compared against that of monolingual toddlers. In addition to evaluating vocabulary size as a compound score, I also studied the phonological and semantic connections between early-learnt words. By studying the structural properties of early vocabulary and how it may affect the acquisition of new words, I aimed to identify heuristics guiding children's vocabulary learning which may help explain differences in vocabulary trajectories between groups.

7.1 Development of vocabulary in toddlers

7.1.1 English vocabulary size

In Chapter 2 of this thesis, I evaluated the vocabulary sizes of bilingual toddlers, comparing them to that of same-age monolingual toddlers. I made use of three methods to calculate vocabulary size – single-language vocabulary size in English, conceptual vocabulary size and total vocabulary size. These three methods tell us about different aspects of bilinguals' vocabulary development. Firstly, single language vocabulary size can tell us about bilinguals' vocabulary development in a specific language of interest. In this case, I looked at the majority language of the community they were growing up in. As expected, I found that bilingual toddlers' vocabulary sizes in English were smaller than the vocabulary sizes of their same-age monolingual peers. This is consistent with previous studies comparing monolingual and bilingual toddlers (Pearson et al., 1993; De Houwer et al., 2014). Additionally, bilingual toddlers' English vocabulary size was predicted by the proportion of English exposure they received in their everyday lives. This finding adds to the literature connecting language exposure and vocabulary growth in bilinguals (Pearson et al., 1997; Hoff et al., 2012; Cattani et al., 2014).

7.1.2 Conceptual vocabulary size

An important consideration when conducting research on bilingual language development is the appropriateness of different measures of vocabulary size to answer the research question. As discussed above, single language vocabulary size is highly dependent on the amount of exposure that the toddler receives in that language. When studying the effect of early vocabulary development in the majority language, for example when predicting school readiness, it may be useful to evaluate single language vocabulary size. But sometimes it is more useful to have a measure that is less subjective to the ratio of language exposure. Conceptual vocabulary size counts any concepts that bilinguals know, regardless of the language it is known in. It presents an estimate of bilinguals' vocabulary that is more representative of bilinguals' language development than single-language vocabulary size. The second comparison that I conducted in Chapter 2 was between bilinguals and monolinguals' conceptual vocabulary size. For bilinguals, concepts were considered to be known if either or both of the translation equivalent pair associated with the concept was known. For English monolinguals, conceptual vocabulary size was equal to their English vocabulary size. I found that bilinguals' conceptual vocabulary size in comprehension was comparable to that of monolinguals. This suggests that learning two languages simultaneously did not have a detrimental effect on bilinguals' toddlers acquisition of new concepts in their vocabulary. This pattern is again consistent with previous research using conceptual vocabulary size (Pearson et al., 1993, 1997). Evaluating vocabulary growth in a way that accounts for growth in both languages may present a better approximation of bilingual toddlers' language ability and later achievement than evaluating only one language.

Related to this, there are both benefits and dangers to treating monolinguals as a "control group" when studying bilinguals' developmental trajectories. One could argue that monolinguals represent learning conditions where only one language is heard, and therefore act as a baseline of performance in the absence of cross-linguistic effects. On the other hand, in addition to linguistic experience, bilinguals often have different cultural experiences as well. While I used a general quantified measure of language exposure for the analyses in Chapter 2, it is important to acknowledge that there is heterogeneity among bilinguals. I reported various aspects of the bilingual toddlers in this sample in Appendix A. The input that bilinguals receive can vary in quantity and also quality. It can also be different in the home and outside the home. Bilingual toddlers' language development has been found to be impacted by the proportion of input received from native speakers of the language (Place & Hoff, 2011; Floccia et al., 2018), parents' proficiency in the language (Hammer et al., 2012; Buac, Gross, & Kaushanskaya, 2014) and the number of different speakers the toddler receives input from (Place & Hoff, 2011). These factors have been theorised to be linked to the quality and diverseness of the input – native speakers and non-native speakers of high proficiency are more likely to produce longer and more complex utterances (Hoff, Core, & Shanks, 2020). Consistent with this, Gámez, Palermo, Perry, and Galindo (2022) found that the lexical diversity of caregivers' input was positively correlated with bilingual toddlers' language development. My recruitment criteria required at least one parent of the child to be a native speaker of the minority language in their community. For families residing in the UK, the minority language would be the AL. The majority of families in the sample (54%) had one parent who was a native speaker of English, and one parent who was a native speaker of the AL. The second most common pattern (33%) was for both parents to be native speakers of the minority language. In Chapter 5, I supplemented the UK sample with data from bilingual families residing in Germany, the Netherlands or Spain. For these families, their minority language would be English. Similar to the UK sample, the majority of families in this sample (60%) had one parent who was a native speaker of English, and one parent who was a native speaker of the AL. Again, the second most common pattern (19%) was for both parents to be native speakers of the minority language (English). The majority of families in both the UK sample and Germany/Netherlands/Spain reported that their child heard proportionally more of their minority language at home, and more of the majority language outside the home. This is consistent with my expectations for the communities the sample was collected in, where there is one dominant majority language used widely in the community and smaller numbers of bilingual speakers of minority languages. While heterogeneity is unavoidable in bilingual research, researchers can help replicability in the field by reporting details of the sample's language environment, identifying both points of overlap and points of divergence.

7.1.3 Total vocabulary size and translation equivalents

In Chapter 2, I also studied the total vocabulary size of bilingual toddlers, which adds together all the words a toddler knows across their two languages. This measure of vocabulary size evaluates how many words the bilingual child knows, regardless of which language the word is known in and whether it overlaps in meaning with another word. As with conceptual vocabulary, monolinguals' total vocabulary size is equal to their English vocabulary size. Bilinguals' total vocabulary size in comprehension was significantly larger than monolingual toddlers' vocabulary size. This pattern is striking when contrasted against their conceptual vocabulary size. Despite both monolinguals and bilinguals having similar conceptual vocabulary sizes, bilinguals knew more words in total. This could be attributed to a high number of doublets, i.e. translation equivalents that are known in both languages, that most of the bilingual toddlers knew. The proportion of doublets among the concepts that toddlers knew was on average 47% for comprehension and 32% in production. As many concepts were known in both languages, bilinguals showed a high total vocabulary size. There are two main possibilities to explain this high proportion of doublets – the first is the words that are easiest to learn in one language are also the words that are easiest to learn in the other. The same words are therefore naturally learnt in the same order in both languages, resulting in incidental overlap. There is a strong possibility that this is at least partially explaining the overlap between languages. However, the large total vocabulary sizes of bilingual toddlers, far surpassing the vocabulary sizes of English monolinguals, lends some support to a second possibility that the shared meaning between words may provide an advantage for learning translation equivalents over learning new concepts. Notably, these two theories are not mutually exclusive, and may in fact be acting in parallel. Unfortunately, the data reported in this thesis is unable to test the validity of these two theories to explain the high proportion of doublets in toddlers' vocabulary. Further research on the acquisition of doublets by bilinguals can tell us more about how overlapping semantic elements between bilinguals' two languages may support the learning of two languages in parallel.

Interestingly, in contrast to total vocabulary size in comprehension, where bilinguals far surpassed monolinguals in the number of CDI words known, total vocabulary size in production was similar across groups. Despite the difference observed between the trajectories in comprehension and production, this does not necessarily suggest that comprehension and production operate using different conceptual systems. Instead, I proposed that this pattern could be linked to maturational constraints that restrict the speed at which toddlers learn to produce new words. These constraints may include developmental milestones of articulatory motor control, where toddlers gradually learn to pronounce phonemes following the specified combinations that form words in the language (Green et al., 2000; Iuzzini-Seigel et al., 2015). According to this theory, it may be easier to produce words that share sounds that toddlers are already familiar with producing. Words with dense phonological neighbourhoods have been found to be produced more consistently and more similar to adult pronunciations than those with lower neighbourhood density (Sosa & Stoel-Gammon, 2012). Researchers have suggested that this may be linked to the frequency of the phonemes' occurrence in toddlers' lexicons. When a newly-learnt word shares phonemes with known words in the lexicon, it may facilitate the formation of strong lexical representations and also facilitate later retrieval. Notably, this facilitatory effect of shared phonemes is likely to apply not only within a language but also across languages. Words that sound similar across languages are likely to be easier to learn, especially if they also share their meaning. The role of cross-linguistic phonological similarity will be discussed in the following section, with a focus on cognates.

7.2 Phonological connections in the lexicon

7.2.1 Calculating phonological similarity

Historically related languages typically feature a relatively large proportion of form-similar or form-identical translation equivalent pairs (known as cognates) than less closely related languages. Cross-linguistic similarity between a foreign word and its translation allows listeners to guess a foreign word's meaning without prior experience with the language. In Chapter 3, I compared two algorithms for phonological similarity against participants' performance in a translation elicitation task where participants had to guess the translation of foreign words without prior knowledge. Participants were most accurate for trials where the presented word sounded identical or very similar to its correct translation.

The decision to use a sample of monolingual participants for this study allowed us to tease apart the effects of phonology and semantics during spoken word recognition. Monolingual participants' performance when listening to foreign words should be sensitive to phonological information provided by the presented word and not its semantic information, due to a lack of prior knowledge about the test language. This allowed me to test the effect of phonological overlap on foreign word recognition without effects of the word's meaning. Another important feature of the experiment design was that the words were presented in auditory format only, without accompanying orthography. By doing this, I aimed to reduce the effect of orthographic overlap on participants' answers. The cognate classifications obtained from the translation elicitation task therefore comprise of cognate pairs where the phonological similarity between words was obvious enough to allow the pair to be identified as cognates with no other cues provided. These cognate classifications are optimal for selecting stimuli to be used in experimental tasks probing the effect of cross-linguistic phonological similarity, as it restricts selection to word pairs that are easily recognisable as cognates.

Both Levenshtein similarity (Levenshtein, 1966; van der Loo, 2014) and ALINE (Kondrak, 2003; Downey et al., 2017), which quantify phonological similarity between two words by calculating the overlap between their phonological transcriptions, performed well in predicting the words that were successfully translated by participants. Trials which had high participant accuracy (>30%) tended to have high scores on Levenshtein similarity and ALINE similarity, while trials with low participant accuracy typically had low scores on both measures. Interestingly, despite ALINE integrating fine-grained phoneme distance into its algorithm, supplementing Levenshtein's all-or-nothing scoring of phoneme overlap, the accuracy of the two measures for classifying word pairs as cognates or non-cognates was largely similar, both very high at around 90%. This is interesting in the context of the translation elicitation task – as participants have no prior knowledge of the test language, it is unsurprising that participants were most sensitive to obvious phonological overlap such as identical phonemes. Participants may have weighted identical phonemes most strongly when making their judgements, with a smaller effect of graded similarity between non-identical phonemes. This possibility is supported by the observation that ALINE did perform slightly better than Levenshtein for classifying cognates. While listeners were sensitive to graded phoneme distance, they weighted identical phonemes most highly when attempting to match a foreign word to a word in their native lexicon.

7.2.2 Competition and lexical frequency

Participants' performance in the translation elicitation task reported in Chapter 3 displayed monolingual participants' ability to use cross-linguistic phonological similarity to guess the translations of foreign words. This strategy was successful for most cognates, but there were some exceptions where participants' accuracy was lower than expected given the degree of similarity between the presented word and its translation. This typically happened when there was strong lexical competition from other similar words in the lexicon. One example was a trial with the presented Catalan word "bol" (English translation "bowl") which is phonologically very similar to English "ball". If participants perceive "bol" to sound more similar to "ball" than "bowl", they may decide to give the false cognate "ball" as their answer.

I explored the effect of lexical competition on participants' foreign word recognition in Chapter 4. I tested a computational model of monolingual spoken word recognition on its ability to model participants' responses in the translation elicitation task. The jTRACE model applies between-word competition when modelling spoken word recognition. In the model, the input word "bol" shares a high similarity with its English translation equivalent "bowl", but also with other English words in the lexicon "ball" and "bow". Presenting the input word "bol" to the jTRACE model activates all of these candidates. All activated candidates in turn apply inhibition on each other as they compete for selection. The degree of inhibition applied on other candidates is proportional to the activation of the source candidate. The presence of strong competitors would therefore reduce the activation of "bol", making it more difficult for "bowl" to be retrieved as the correct translation. In contrast, Spanish input word "pinguino", which also has high similarity its English translation "penguin", has no close competitors in the English lexicon. Due to low phonological similarity, other candidate words would be activated weakly in response to "pinguino", and therefore be unable to apply strong inhibition to reduce the activation of the candidate word "penguin". This allows easier and unambiguous selection of "penguin" as the final candidate. In Chapter 4, I showed that the jTRACE model, which integrates this element of lexical competition, performed better than both Levenshtein similarity and ALINE which only account for phonological overlap. This highlights how other words in the lexicon, not only the target word, can affect spoken word recognition.

7.2.3 Facilitation from cognates

Phonological connections with other words in the lexicon can affect the acquisition of new words as well. When learners are learning a second language, cognates can facilitate learning while false cognates can inhibit word learning and word recognition. In Chapter 5, I theorised that bilingual toddlers can make use of cross-linguistic phonological similarity between translation equivalents to learn new words more easily. The facilitatory effect of cognates was supported by findings in Chapter 5, where cognates were found to be learnt earlier than non-cognates by bilingual toddlers. Toddlers are able to identify regularities both within a language and between languages, and make use of this systematicity. Despite being in the early stages of learning and having no explicit taught knowledge of cross-linguistic similarity, toddlers are able to make use of cognates to aid their language learning.

Theories of the cognate facilitation effect emphasise the parallel activation between bilinguals' two languages, where the shared phonological information between cognates prompts higher activation as the two words feed activation to each other. In the case of bilingual word learning, hearing a novel word should activate phonologically-similar words in the lexicon, as posited by the jTRACE model simulations discussed in the previous section. If a Spanish-English bilingual child hears the novel Spanish word "pinguino" and they already know the word "penguin" in English, the word "penguin" should be automatically activated through its phonological similarity with the novel word. As "penguin" has become more strongly activated than other English words, it should be easier for the child to match "pinguino" and "penguin" to the same concept. The shared activation may also help strengthen the lexical representations of the newly-acquired word. In contrast, while a non-cognate novel word, for example "perro", would also activate phonologically-similar candidates in the lexicon like "pear", it would not facilitate the acquisition of the novel word due to the low activation of the correct translation "dog". Importantly, the theory holds the assumption that for cognates to show an advantage in word learning, the translation equivalent of the novel word must already be present in the learner's lexicon. This is supported by findings in Chapter 5 that the cognate facilitation effect was modulated by existing word knowledge. When one word of a translation equivalent pair was already known by the toddler, the other word in the pair was more likely to be learnt if they were cognates than if they were non-cognates. In contrast, if neither word of a translation equivalent pair was known, there was no difference in the likelihood of learning between cognates and non-cognates. This effect of existing knowledge on the cognate facilitation effect once again highlights the role of toddlers' existing lexicon on the acquisition of new words.

7.3 Semantic connections in the lexicon

As discussed in the previous section, the learning of new words can be helped and hindered by connections to known words. So far I have discussed this effect in relation to phonological connections between words, and in the following section I will move on to discuss it in relation to semantic connections. Words in the lexicon have semantic relationships with each other through their shared features, shared taxonomic category, associative relationships, or any combination of the above.

In Chapter 6, I started by investigating the factors predicting the acquisition order of early words learnt by monolingual toddlers. This was tested on the group level, where a word was considered to be acquired by a certain age (in months) if at least 50% of toddlers in that age group were reported by their parents to understand (for age-of-acquisition in comprehension) or produce (for age-of-acquisition in production) the word. This provided a relative scoring of word difficulty, with earlier-learnt words considered easier to learn and later-learnt words considered more difficult. A key finding of this study was that different patterns were observed for the four word classes analysed. Firstly, the acquisition order of nouns was predicted by both frequency and associative diversity (as operationalised using associative in-degree), with unique variance contributed by each of these two predictors. This suggested that nouns were learnt earlier if they occurred more frequently in the environment and if they occurred across different contexts. A noun's frequency of occurrence is likely to correlate with the frequency of the referent object's appearance in the toddlers' vicinity, particularly for nouns known by toddlers which are mostly concrete and refer to objects in their daily lives (e.g.

animals, toys, furniture). Higher frequency of a word could therefore facilitate learning of the word-referent mapping. Associative diversity may also help the disambiguation of word meaning, allowing toddlers to match the novel word to the correct object over repeated occurrences in different contexts. This is the basis of cross-situational learning theories (Pinker, 1994; Gleitman, 1990; Siskind, 1996; Frank, Goodman, & Tenenbaum, 2009; Blythe, Smith, & Smith, 2010; K. Smith, Smith, & Blythe, 2011) – if a dog appears together with a cat in one scene and with a rabbit in another, and the child hears the word "dog" across both scenes, this can help them disambiguate the referent of "dog" from the possible choices of a dog, a cat or a rabbit. When target word-referent pairs occur in a wide variety of contexts, learners are more successful at inferring word-referent mappings than when the target pairs only occur in limited contexts (Kachergis et al., 2009; Suanda et al., 2014). Theories on how this inference process occurs include associative learning mechanisms (K. Smith et al., 2011; McMurray et al., 2012) or Bayesian hypothesis testing (Xu & Tenenbaum, 2007).

Associative diversity also predicted the acquisition order of function words. Given the abstract nature of function words, associative diversity may help the inference of meaning by providing cues about the meaning through its relationships with other words. Unlike with nouns, frequency was a very poor predictor for the acquisition of function words, suggesting that frequent repetitions of the words does not in itself support learning, but instead context of occurrence is more important for this abstract class of words. This pattern of results is consistent with the assumptions of a model by Hidaka (2013), which proposed function words are learnt through inference from the sentential context and are relatively insensitive to effects of lexical frequency.

Unexpectedly, neither frequency nor associative diversity significantly predicted the acquisition order of verbs and adjectives. The majority of verbs and adjectives in the CDI are relatively concrete, for example action verbs like “jump” and colour words like “white”. The process of learning verbs and adjectives that describe physical traits is made more difficult by the way these words naturally co-occur in the environment with nouns, making their meaning more difficult to infer without further cues. When their parent refers to a jumping white rabbit, a child would need to infer the respective meanings of the verb “jump”, the adjective “white” and the noun “rabbit”. This is further complicated by the multitude of other possible referents in the scene, for example the body parts of the rabbit. This problem has been referred to as referential ambiguity or the gavagai problem (Quine, 1960). Word learning biases such as a bias towards assuming that a novel word refers to a novel whole object (as opposed to a specific part of the object or other more abstract referents) (Markman & Wachtel, 1988) and a shape bias when extending novel nouns to new instances (Landau, Smith, & Jones, 1988) have been proposed to support learning of nouns, which is consistent with the pattern that toddlers’ earliest-learnt words are mostly concrete nouns. With these early word learning heuristics biasing learners towards interpreting novel words as nouns, learners need to make use of additional cues to correctly disambiguate the meaning of verbs and adjectives from other possible referents in the environment. Such cues may include syntactic bootstrapping, where children can use the syntactic structure of a sentence to infer the meaning of a verb (Landau & Gleitman, 2009; L. Naigles, 1990). It has been suggested that verb learning is supported by predictability of the context it occurs in, as opposed to associative diversity as explored in this study. The differences in the predictive power of frequency and associative diversity observed between nouns, verbs, adjectives and function words highlight how different types of cues may be needed to support learning of different word classes.

In the latter half of Chapter 6, I studied the structural properties of semantic networks built from the vocabulary known by toddlers using semantic network analyses. I built a network for each child's vocabulary using associative connections obtained from the Small World of Word free association database (De Deyne et al., 2019). Words were connected by an edge if one of the words had been given as a response when the other was a cue. The presence of an edge therefore signifies that those two words are associated. The vocabulary networks were compared against random networks with the same number of nodes. Two types of random networks were constructed for each toddlers' vocabulary network – an Erdős–Rényi (ER) network and a random acquisition network. An ER network has the same number of nodes and same number of edges as the vocabulary network but with the edges randomly assigned in the network. An ER network represents the expected level of connectivity if there were no systematic patterns in word learning order nor the learning environment. The second type, random acquisition networks, are built by randomly selecting words from the CDI equal to the number of words in the toddlers' vocabulary. The edges in the random acquisition network are established using the SWOW free associations, just like the vocabulary networks. Random acquisition networks represent a level of connectivity when there is a systematic pattern in the learning environment but words are not learnt in a systematic order. By comparing vocabulary networks against random networks, I can investigate whether vocabulary development follows a systematic pattern that can be characterised by associative links. I used three metrics to quantify the connectivity of semantic networks – clustering coefficient, average in-degree and geodesic distance. A higher clustering coefficient, higher average in-degree and shorter geodesic distance as compared to random networks would suggest that toddlers' vocabulary networks have high associative connectivity.

Compared to random networks, toddlers' vocabularies showed higher connectivity, suggesting that toddlers were sensitive to patterns in the language environment. The previously-discussed findings that the acquisition of nouns and function words were predicted by associative in-degree is consistent with this pattern – toddlers learn words that are highly connected in the learning environment earlier than less connected words, and this results in high connectivity in the early lexicon. Small differences in clustering coefficient was discovered between toddlers whose vocabulary sizes fall in the lowest percentiles of their age group and those in the highest percentile. This difference was only found for clustering coefficient and not the other two measures of network connectivity, suggesting that the effect is small. The vocabulary of toddlers' from all groups showed higher connectivity compared to random networks, suggesting that the vocabulary development of toddlers from these different groups still followed a similar pattern that is sensitive to associative links in the language.

A clearer difference was observed between monolingual and bilingual toddlers, with bilinguals showing higher connectivity across all 3 tested measures. This pattern of results suggests that bilinguals' word learning trajectories follow associative patterns even more than monolinguals. This is interesting to consider in relation to cross-linguistic influences between languages as discussed in previous chapter. Many associative links are expected to be consistent across languages (e.g. body part associations like “finger” and “hand”) while a subset of associations may be culture-specific (e.g. “duck” and “bathtub”). This mix of overlapping and non-overlapping associations across languages may have led to word learning patterns not fully captured by the present study, as the analysed semantic networks were built using associations collected from English speakers only. It is also not fully clear whether to represent semantic networks as conceptual (1 node per concept, regardless of language) or language-specific (2 nodes per concept, one for each translation equivalent). Further research in this field can help broaden our understanding of cross-linguistic semantic connections and its possible effects on bilingual toddlers’ word learning.

7.4 Effect of the COVID-19 pandemic on studies

The COVID-19 pandemic had an unfortunate impact on the direction of my DPhil projects. Two in-lab experimental studies were originally planned to be included in this thesis, but due to a complete inability to run in-person studies during the lockdowns and a persisting low participant recruitment rate even after restrictions were eased, they could not be completed in time. The first experimental study was designed to investigate the effects of non-selective bilingual lexical access on toddlers’ spoken word recognition, with a focus on cognates. It would have supplemented the discussions of cross-linguistic similarity in this thesis, forming a bridge between the studies conducted on toddlers' word learning and adults' spoken word recognition. The second experimental study would have investigate the effect of increasing vocabulary size on the formation and strengthening of semantic connections between words in toddlers' lexicon. It would have fit into the theme of semantic connections between words in the latter half of this thesis.

In response the lockdown restrictions, I shifted resources towards questionnaires and computerised behavioural studies. The final set of studies reported in this thesis were conducted with data collected online. Participants and families were reached remotely via social media, email and online recruitment platforms. In relation to this, it is important to be aware of the characteristics of the collected sample. This is particularly relevant for all data collected from or about infant participants. The CDI data analysed in Chapters 2 5 was collected exclusively between 2020-2022, the earlier part of this period corresponding to national lockdowns in response to the COVID-19 pandemic. The restrictions associated to these lockdowns resulted in many children having less opportunity to engage in social activities outside the home than they likely would have otherwise. There were several anecdotal comments from parents that their child was unable to continue attending nursery as per their normal schedule due to lockdown nursery closures. To minimise the possibility of confounds, I applied some checks to the data. I compared the monolinguals data from the reported sample (collected between 2020 and 2022) against older data collected before 2020. I found no significant differences between the vocabulary sizes of the two samples, suggesting that there was no obvious effect of the lockdowns on toddlers' vocabulary growth. This pattern is similar to that found by Sperber, Hart, Troller-Renfree, Watts, and

Noble (2022), who found no significant relationship between the amount of time spent in pandemic lockdown conditions and language outcomes at 24 months old. There was a small decrease in language development measures at 12 months old but this was not significant. Another study, by Kartushina et al. (2021), actually found higher vocabulary for toddlers growing up during the recent lockdowns than age-matched toddlers whose data was collected prior to 2020, which the authors linked to more time spent with parents during lockdown. For all analyses comparing monolingual and bilinguals in this thesis, both samples were collected over the 2020–2022 time period, allowing them to be compared directly. While it is possible that the bilingual families in the sample faced different circumstances from the monolingual families sampled at the same time, I do not have sufficient data to investigate this speculation. I make the assumption that the lockdowns affected both monolingual and bilingual families in similar ways.

One last element to be considered in relation to the pandemic is whether bilingual toddlers' exposure to their two languages may have been different compared to times without social and travel restrictions. The common occurrence of parents working from home during the lockdown may have resulted in more exposure to the minority language for the child. On the flip side, travel restrictions meant that some families may not have been able to travel to spend time with extended family members. The bilingual child may have received less opportunity to interact with additional speakers of their AL aside from their parents. For the majority language, the child may have received less exposure to the majority language due to not attending nursery and playgroups. These elements are expected to be captured by the language exposure questionnaire parents answered together with the CDIs, which asks parents to report their child's overall language exposure and home language exposure. In-depth evaluations of bilingual toddlers' language exposure before, during and after the COVID-19 pandemic can shed light on the impact of the pandemic on bilingual toddlers' language environment and subsequent vocabulary development.

7.5 Future directions

7.5.1 Words with related but non-identical meanings

In this thesis, I discussed the effects of cognates and false cognates on word learning and recognition. Among word pairs with cross-linguistic phonological similarity, there is a subset of words with more complicated relationships. These word pairs share phonological overlap and the same meaning, but the translation has a more frequently-used competitor. One example is the German noun "Blom" and the English noun "bloom". While the noun "bloom" can be used as the translation of "Blom", it is much more common in modern English usage to translate the word as "flower". A German-English bilingual using the word "bloom" as a translation for "Blom" would not be incorrect, but it may be viewed as an unusual choice of words for a monolingual speaker of English. Word pairs such as this fall into an intersection between cognates and false cognates. Another ambiguous case would be phonologically-similar word pairs which are semantically-related but not translation equivalents. The English adjective "lunar" (used to refer to things related to the moon) shares etymology and phonological overlap with the Spanish noun “luna” (moon). However, the English translation for “luna” is “moon” (non-cognates). Nevertheless, English speakers learning Spanish could make use of their knowledge of English “lunar” to guess the meaning of Spanish “luna”. While this may not be a strategy available to toddlers (“lunar” is typically a late-learnt word), adult second language speakers can utilise these adjacent similarities to help their learning. Future research on how these ambiguous word pairs may affect word learning and word retrieval can tell us more about how learners make use of semantic and phonological connections in their lexicon.

7.5.2 Assumptions of corpus-derived statistics

The use of corpus-derived statistics of frequency and contextual diversity makes strong assumptions about the linguistic environment systematically reflecting the physical learning environment. Studies that collect both video and audio data on children’s day-to-day activities can provide us with crucial supplementary information on the physical environment that accompanies everyday speech. Systematic exploration of the rich data offered by such studies can help us reach past the abstractness of purely linguistic data to identify what these derived statistics actually represent in a real-world learning environment. Further work also needs to be done for operationalising associative diversity for infant research. The large free association databases typically used in semantic network studies are comprised of free association data from adult participants. When calculating associative diversity for the study in Chapter 6, I used the subset of concepts present in the CDI, thus restricting the associations analysed to those between early-learnt words only. Nevertheless, while I expect there to be many commonalities between adult and infant associations (e.g. "table" and "chair" which co-occur in the physical environment), there may be some associations made by adults that are not made by children. One example is a link between "cake" and "walk" from the informal phrase "cakewalk", which is unlikely to be known by toddlers. Conversely, they may also be some associations that are stronger for infants than adults, for example an association between "mother" and "milk". Infant research on semantic connections between words would be benefit from more infant-focused databases of associations.

7.6 Concluding Remarks

In this thesis, I covered topics of vocabulary growth, phonological connections and semantic connections, focusing on the rich interaction between words in a toddlers' lexicon. I showed how toddlers are sensitive to cross-linguistic similarity in the form of cognates. Toddlers are able to make use of this similarity to learn cognates faster than non-cognates. Additionally, toddlers show a bias towards learning translation equivalents across their languages. The shared meaning between translation equivalents may help the acquisition of words. Toddlers are also sensitive to the semantic relationship between words in their learning environment, preferentially learning high frequency and highly connected words first. Taken together, the studies in this thesis show that bilingual toddlers are able to capitalise on systematicity in their learning environment in the form of both phonological connections and semantic connections to facilitate learning two languages in parallel.

Acknowledgements

Ethics references: R60939/RE009 (bilingual data); R63335/RE007 (monolingual CDI data). Informed consent was obtained for studies with human participants.

Thanks to Irina Lepadatu and Dr Nicola A. Gillen for helping collect the bilingual and monolingual CDI data reported in this thesis.

Thanks to Gonzalo Garcia-Castro for helping with data collection in the translation elicitation task reported in this thesis.

Thanks to Prof Kim Plunkett for supervision of this DPhil project.

Thanks to associates who helped with adapting the translations of the Oxford CDI used in this thesis:

- Dutch: Prof Paula Fikkert
- German: Dr Marlene Spangenberg
- French: Prof Caroline Floccia
- Italian: Dr Davide Volpi

- Polish: Zuzanna Patryas
- Portuguese: Andreia Correia Brecha
- Spanish: Prof Nuria Sebastian-Galles, Daniela Avila-Varela and Gonzalo Garcia-Castro

Thanks to Dr Janette Chow for guidance provided on analysing corpus data and semantic networks.

Thanks to Prof James Magnuson for advice regarding jTRACE.

Huge thanks to participants and families who contributed to the studies in this thesis.

Funding: Studies in this thesis were supported by the Economic and Social Research Council grant CQR01830 to Kim Plunkett and Nuria Sebastian-Galles. Serene Siow was employed as a Research Assistant under this grant for partial duration of her DPhil.

Re-use of published material in thesis

Written permission to include material from the paper in this thesis has been acquired from all authors listed below.

Chapter 2: Comparing bilingual toddlers' vocabulary size with monolingual toddlers' vocabulary size

Siow, S., Lepadatu, I., Gillen, N., & Plunkett, K. (2022). UK bilingual toddlers show a lag in vocabulary size relative to monolinguals in both comprehension and production. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 44, No. 44).

Student's contribution to paper: collected data (with help from IL and NG), analysed data, interpreted results, wrote manuscript

Chapter 5: Cross-linguistic similarity between words facilitates bilingual toddlers' vocabulary trajectories

Siow, S., Gillen, N. A., Lepadatu, I., Avila-Varela, D., Garcia-Castro, G., Sebastian-Galles, N., & Plunkett, K. (2022). The effect of cognates on bilingual infant vocabulary trajectories: a study using bilingual CDIs of English and one additional language. In Proceedings of the 46th Annual Boston University Conference on Language Development. Cascadilla Press.

Student's contribution to paper: collected data (with help from IL and NG), analysed data, interpreted results, wrote manuscript

Chapter 6: Exploring the variable effects of frequency and semantic diversity as predictors for a word's ease of acquisition in different word classes

Siow, S., & Plunkett, K. (2021). Exploring the variable effects of frequency and semantic diversity as predictors for a word's ease of acquisition in different word classes. Proceedings of the Annual Meeting of the Cognitive Science Society, 43. Retrieved from https://escholarship.org/uc/item/83t6n1rq

Student's contribution to paper: analysed data, interpreted results, wrote manuscript

Appendix A

Participant demographics for CDI analyses (bilingual sample)

A.1 Questionnaire Items

A.1.1 Demographics questionnaire

1. Child's date of birth
2. Your child's gender
3. Please, could you state your postcode (optional)
4. Mother / Parent 1's highest educational qualification (optional)
5. Father / Parent 2's highest educational qualification (optional)

6. Has your child ever had any hearing problems, including glue ear?
7. Was your child born premature? If yes, how many weeks early?
8. What was your child's weight at birth (in grams)?

A.1.2 Language Exposure Questionnaire

Adapted from the bilingual language exposure questionnaire developed by Bosch and Sebastián-Gallés (2001).

Example questionnaire for bilingual English-Spanish family

This questionnaire will ask you about the languages your child is exposed to in their day-to-day routine.

Parent-estimated percentage of exposure

1. On average, what percentages of these languages does your child hear? Think about the languages your child hears at home (parents, siblings, grandparents, other household members) and outside the home (nursery, playgrounds, shopping, holidays, friends), considering the frequency of interaction and words spoken directly to your child. (sum to 100%)
    (a) Spanish
    (b) English
    (c) Other(s) [*FREE TEXT BOX*]

2. What percentages of these languages does your child hear just at home? Think specifically about the languages each parent, siblings and other household members speak to your child. (sum to 100%)
    (a) Spanish
    (b) English
    (c) Other(s) [*FREE TEXT BOX*]

Mother / Carer 1

This first section should be completed by the child's mother, or where not applicable, the primary carer.

1. What is your native language?
    (a) Spanish
    (b) English
    (c) Bilingual Spanish and English
    (d) Other(s) [*FREE TEXT BOX*]
2. How would you judge your spoken proficiency in these languages? (Native = 10; Do not speak language = 0)
    (a) Spanish
    (b) English
    (c) Other(s) [*FREE TEXT BOX*]

3. Which of the following best describes the language(s) you use when speaking to your child?
    (a) I always (or almost always) speak Spanish to my child
    (b) I always (or almost always) speak English to my child
    (c) I regularly speak both Spanish and English to my child
    (d) I regularly speak a mix of Spanish / English / other languages to my child

Father / Carer 2

The following subsection should be completed by the child's father or second carer. (please continue to next subsection if not applicable)

1. What is your native language?
    (a) Spanish
    (b) English
    (c) Bilingual Spanish and English
    (d) Other(s) [*FREE TEXT BOX*]
2. How would you judge your spoken proficiency in these languages? (Native = 10; Do not speak language = 0)
    (a) Spanish
    (b) English

    (c) Other(s) [*FREE TEXT BOX*]

3. Which of the following best describes the language(s) you use when speaking to your child?
    (a) I always (or almost always) speak Spanish to my child
    (b) I always (or almost always) speak English to my child
    (c) I regularly speak both Spanish and English to my child
    (d) I regularly speak a mix of Spanish / English / other languages to my child

Siblings

1. Within the family, is the child the first / second / third born child?
    (a) First
    (b) Second
    (c) Third
    (d) Fourth
    (e) Other [*FREE TEXT BOX*]

Nursery

1. At what age did your child start attending daycare / nursery? (age in months)
2. (a) Always (or almost always) Spanish

(b) Always (or almost always) English

(c) Both Spanish and English regularly

(d) A regular mix of Spanish / English / other languages

Community

1. Cumulatively, approximately how many months has your child spent in a country / community where Spanish is the primary language? (e.g. visits longer than a week, previous residence in Spanish-speaking country)

Additional information

1. If you would like to share any additional details about life events that may have affected your child's language environment, please write them below.

A.2 Participant demographics

I collected data online from bilingual families who speak English and one other language in their daily lives. The final sample after exclusions were 779 participants (N female = 399; N male = 379; N other = 1), some which contributed longitudinal data. The total number of questionnaires collected inclusive of longitudinal entries was 1185. The majority of the participants were collected from the UK (N = 683), with smaller samples from Germany (N = 48), the Netherlands (N = 28) and Spain (N = 80).

Table A.1 shows the sample size, mean age, mean overall English exposure and mean home English exposure, split across the different language pairs and countries they were recruited from.

Table A.1: Number of participants split by language and country, with mean age in months, mean overall English exposure (%) and mean home English exposure (%) at time of first questionnaire. Standard deviation presented in brackets.

language	country	N	age	overall eng	home eng
Dutch	Netherlands	28	25.4 (6.7)	53.8 (14.3)	69.6 (21.6)
Dutch	UK	34	25.1 (6.7)	61.9 (16.6)	48.2 (21.8)
French	UK	98	23.3 (7)	61 (17.3)	40.1 (22.7)
German	Germany	48	25.4 (7)	43.6 (18)	65.1 (15.9)
German	UK	47	25.7 (6.3)	60.7 (15.6)	48.6 (22.3)
Italian	UK	105	23 (6.4)	54.5 (17.5)	30.5 (24.1)
Polish	UK	99	23.3 (7.3)	50.3 (19.1)	33.8 (21.5)
Portuguese	UK	58	24.5 (6.3)	50.7 (19)	28.6 (22.6)
Spanish	Spain	80	24.7 (6.9)	50.9 (17.6)	63.8 (23.6)
Spanish	UK	182	23.7 (7)	51.1 (17.4)	37.2 (21.3)

Note. age = Mean age in months; overall eng = mean overall English exposure (%); home eng = mean home English exposure (%)

Parent-estimated overall language exposure

The mean overall English language exposure for the UK sample was 54.3 (SD = 17.8). The mean for the Germany, Netherlands and Spain sample combined was 47.0 (SD = 17.3). The distribution of overall English language exposure, split by region, is illustrated in a histogram in Figure A.1.

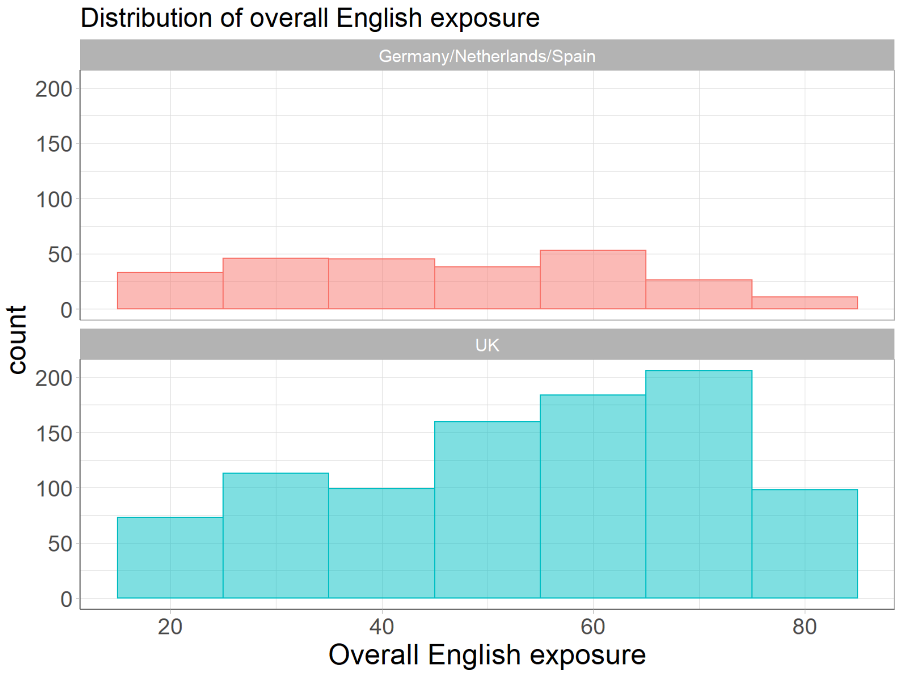

Figure A.1: Histogram of parent-reported overall English language exposure percentage, split by region.

Parent-estimated home language exposure

The mean home English language exposure for the UK sample was 37.4 (SD = 22.5). The mean for the Germany, Netherlands and Spain sample combined was 64.0 (SD

= 20.6). The distribution of home English language exposure, split by region, is illustrated in a histogram in Figure A.2.

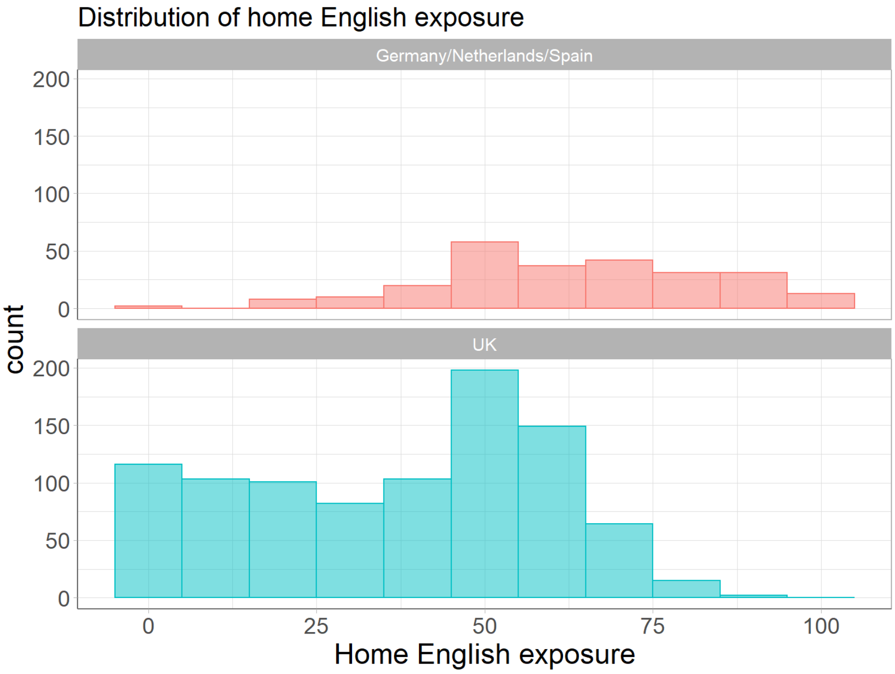

Figure A.2: Histogram of parent-reported home English language exposure percentage, split by region.

Parent native language

The distribution of parents' native languages from the UK sample are shown in Table A.2. The distribution of parents' native languages from the Germany/Netherlands/Spain sample are shown in Table A.3.

Table A.2: Distribution of each parent's native language (UK sample)

native language	N	%
Mother		
AL	535	72.4
English	49	12.8
Bilingual English and AL	32	11.5
Other(s)	7	1.12
Father		
AL	255	40.9
English	322	51.7
Bilingual English and AL	17	2.73
Other(s)	29	4.65

Table A.3: Distribution of each parent's native language (Germany/Netherlands/-Spain sample)

native language	N	%
Mother		
AL	20	12.8
English	113	85.9
Bilingual English and AL	32	11.5
Other(s)	5	3.2
Father		
AL	86	55.1
English	56	35.9
Bilingual English and AL	12	7.69
Other(s)	2	1.28

From the above information on parents' native language, I derived the combinations of parents' native languages in the this sample, listed in Table A.4.

Table A.4: Distribution of parent native language combinations

native language	N	%
UK		
1 AL native 1 Eng native	339	54.4
Both AL native	205	32.9
Both English native	1	0.161
1 Bilingual 1 AL native	15	2.41
1 Bilingual 1 Eng native	25	4.01
Includes third language	38	6.10
Germany/Netherlands/Spain		
1 AL native 1 Eng native	94	60.3
Both AL native	0	0
Both English native	30	19.2
1 Bilingual 1 AL native	9	5.77
1 Bilingual 1 Eng native	11	7.05
Includes third language	12	7.69

Parent language proficiency

Parents' self-rated language proficiency are shown in Table A.5.

Parent language use

The distribution of parents' usual language use when speaking to child, from the UK sample are shown in Table A.6. The distribution of parents' usual language use when speaking to child, from the Germany/Netherlands/Spain sample are shown in Table A.7.

Table A.5: Distribution of each parent’s self-rated language proficiency

nativeness	Eng proficiency	AL proficiency
UK		
Mother		
(native English)	9.98 (0.143)	5.06 (2.61)
(native AL)	8.08 (1.09)	9.98 (0.155)
(bilingual)	9.47 (1.14)	9 (1.67)
(native other)	8.86 (1.21)	9 (1.91)
Father		
(native English)	9.98 (0.184)	2.74 (2.24)
(native AL)	7.86 (1.38)	9.97 (0.233)
(bilingual)	9.41 (1.50)	8.59 (2.81)
(native other)	8.69 (1.11)	4.79 (3.86)
Germany/Netherlands/Spain		
Mother		
(native English)	9.99 (0.0941)	6.39 (2.33)
(native AL)	7.75 (1.41)	10 (0)
(bilingual)	9.72 (0.752)	9.33 (1.41)
(native other)	9.6 (0.548)	8 (2.92)
Father		
(native English)	9.88 (0.689)	4.71 (2.55)
(native AL)	6.90 (2.41)	10 (0)
(bilingual)	9.25 (1.14)	9.67 (0.651)
(native other)	8 (1.41)	9.5 (0.707)

Table A.6: Distribution of each parent’s language use to child (UK sample)

native language	N	%
Mother		
Always AL	404	64.8
Always English	46	7.38
Both English and AL regularly	153	24.6
Mix of English, AL and other languages	20	3.21
Father		
Always AL	222	35.6
Always English	304	48.8
Both English and AL regularly	79	12.7
Mix of English, AL and other languages	18	2.89

Table A.7: Distribution of each parent’s language use to child (Germany/Netherlands/Spain sample)

native language	N	%
Mother		
Always AL	12	7.69
Always English	115	73.7
Both English and AL regularly	25	16.0
Mix of English, AL and other languages	4	2.56
Father		
Always AL	66	42.3
Always English	54	34.6
Both English and AL regularly	29	18.6
Mix of English, AL and other languages	7	4.49

Parent education level

The distributions of parents' education level are shown in Table A.8.

Table A.8: Distribution of each parent's education level

highest education level	N	%
Mother		
University degree or equivalent	674	86.5
Left school at 18 with A-Levels or equivalent	73	9.37
Left school at 16 with GCSE or equivalent	25	3.21
No qualifications	4	0.513
Not reported	3	0.385
Father		
University degree or equivalent	566	72.7
Left school at 18 with A-Levels or equivalent	114	14.6
Left school at 16 with GCSE or equivalent	84	10.8
No qualifications	8	1.03
Not reported	7	0.899

Birth order

Table A.9: Distribution of child's birth order among siblings

birth order	N	%
First	542	69.6
Second	187	24.0
Third	39	5.01
Fourth	4	0.51
Fifth	1	0.13
Twins	6	0.77

Community and time abroad

Out of the UK sample, 43.5% have spent time immersed in a country where their home language (i.e. the AL) is spoken in the community. The other 56.5% were reported to not have spent time (or less than a month). Of the toddlers who have language immersion experience in the UK sample, the average time spent is 3.43 months (SD = 3.89). When converted into the percentage of their age at the time of reporting, the mean percentage of their lives spent immersed was 14.5% (SD = 15.9).

Out of the Germany/Netherlands/Spain sample, 28.2% have spent time immersed in a country where their home language (i.e. English) is spoken in the community. The other 71.8% were reported to not have spent time (or less than a month). Of the toddlers who have language immersion experience in this sample, the average time spent is 4.30 months (SD = 5.54). When converted into the percentage of their age at the time of reporting, the mean percentage of their lives spent immersed was 16.5% (SD = 22.0).

Nursery attendance

Out of the sample, 68.1% were reported to have started nursery at the time of the questionnaire. The mean age of starting nursery was 13.0 months (SD = 6.33) among the subset who had started nursery.

Appendix B

Oxford Communicative Development Inventory

Table B.1: Oxford CDI words and categories.

	word	category
1	baa baa (sheep)	Sounds
2	choo choo (train)	Sounds
3	cockadoodledoo (rooster)	Sounds
4	grr (growl)	Sounds
5	meow (cat)	Sounds
6	moo (cow)	Sounds
7	ouch	Sounds
8	quack (duck)	Sounds
9	uh oh	Sounds

10	vroom (car)	Sounds
11	woof (dog)	Sounds
12	yum	Sounds
13	animal	Animals
14	bear	Animals
15	bee	Animals
16	bird	Animals
17	bunny / rabbit	Animals
18	butterfly	Animals
19	cat	Animals
20	chicken	Animals
21	cow	Animals
22	deer	Animals
23	dog	Animals
24	donkey	Animals
25	duck	Animals
26	elephant	Animals
27	fish	Animals
28	frog	Animals
29	giraffe	Animals
30	goose	Animals
31	horse	Animals
32	kitten	Animals

33	lamb	Animals
34	lion	Animals
35	monkey	Animals
36	mouse	Animals
37	owl	Animals
38	penguin	Animals
39	pig	Animals
40	pony	Animals
41	puppy	Animals
42	sheep	Animals
43	spider	Animals
44	squirrel	Animals
45	tiger	Animals
46	turkey	Animals
47	turtle	Animals
48	aeroplane / plane	Vehicles
49	bicycle / bike	Vehicles
50	boat	Vehicles
51	lorry / truck	Vehicles
52	motor-bike	Vehicles
53	bus	Vehicles
54	car	Vehicles
55	fire engine	Vehicles

56	pushchair / buggy	Vehicles
57	train	Vehicles
58	ball	Toys
59	balloon	Toys
60	block / brick	Toys
61	book	Toys
62	bubble	Toys
63	doll	Toys
64	pen	Toys
65	teddy bear	Toys
66	toy	Toys
67	apple	Food and Drink
68	banana	Food and Drink
69	biscuit	Food and Drink
70	bread	Food and Drink
71	butter	Food and Drink
72	cake	Food and Drink
73	carrot	Food and Drink
74	cereal	Food and Drink
75	cheese	Food and Drink
76	chicken	Food and Drink
77	chips	Food and Drink

78	coffee	Food and Drink
79	drink	Food and Drink
80	egg	Food and Drink
81	fish	Food and Drink
82	food	Food and Drink
83	ice cream	Food and Drink
84	jam	Food and Drink
85	juice	Food and Drink
86	meat	Food and Drink
87	milk	Food and Drink
88	orange	Food and Drink
89	pasta / spaghetti	Food and Drink
90	peas	Food and Drink
91	pizza	Food and Drink
92	sweets	Food and Drink
93	tea	Food and Drink
94	toast	Food and Drink
95	water	Food and Drink
96	arm	Body Parts
97	belly button / tummy button	Body Parts
98	cheek	Body Parts
99	ear	Body Parts
100	eye	Body Parts

101	face	Body Parts
102	finger	Body Parts
103	foot	Body Parts
104	tongue	Body Parts
105	tooth	Body Parts
106	hair	Body Parts
107	hand	Body Parts
108	head	Body Parts
109	knee	Body Parts
110	leg	Body Parts
111	nail	Body Parts
112	nose	Body Parts
113	toe	Body Parts
114	tummy	Body Parts
115	mouth	Body Parts
116	bib	Clothes
117	boot(s)	Clothes
118	button	Clothes
119	coat	Clothes
120	dress	Clothes
121	glasses / specs	Clothes
122	hat	Clothes
123	jacket	Clothes

124	jeans	Clothes
125	jumper /sweater	Clothes
126	nappy	Clothes
127	necklace	Clothes
128	pyjamas	Clothes
129	shirt	Clothes
130	shoe	Clothes
131	shorts	Clothes
132	sock	Clothes
133	trousers	Clothes
134	zip	Clothes

135	bath / bathtub	Furniture and Rooms
136	bathroom	Furniture and Rooms
137	bed	Furniture and Rooms
138	bedroom	Furniture and Rooms
139	chair	Furniture and Rooms
140	cooker / stove /oven	Furniture and Rooms
141	cot	Furniture and Rooms
142	door	Furniture and Rooms
143	drawer	Furniture and Rooms
144	garage	Furniture and Rooms
145	high chair	Furniture and Rooms
146	kitchen	Furniture and Rooms

147	living room	Furniture and Rooms
148	play pen	Furniture and Rooms
149	potty	Furniture and Rooms
150	refrigerator / fridge	Furniture and Rooms
151	rocking chair	Furniture and Rooms
152	settee /sofa	Furniture and Rooms
153	sink	Furniture and Rooms
154	stairs	Furniture and Rooms
155	table	Furniture and Rooms
156	TV / television	Furniture and Rooms
157	window	Furniture and Rooms
158	beach	Outside
159	bucket	Outside
160	church	Outside
161	flower	Outside
162	garden	Outside
163	house	Outside
164	moon	Outside
165	sky	Outside
166	slide	Outside
167	snow	Outside
168	spade	Outside
169	star	Outside

170	stone	Outside
171	sun	Outside
172	outside	Outside
173	park	Outside
174	party	Outside
175	pool	Outside
176	rain	Outside
177	school	Outside
178	shop	Outside
179	swing	Outside
180	tree	Outside
181	wall	Outside
182	water	Outside
183	work	Outside
184	zoo	Outside
185	bin	Household items
186	blanket	Household items
187	bottle	Household items
188	bowl	Household items
189	box	Household items
190	broom	Household items
191	brush	Household items
192	clock	Household items

193	comb	Household items
194	cup	Household items
195	dish	Household items
196	dummy	Household items
197	fork	Household items
198	glass	Household items
199	hammer	Household items
200	hoover /vacuum	Household items
201	jug	Household items
202	key	Household items
203	lamp	Household items
204	light	Household items
205	medicine	Household items
206	money	Household items
207	mug	Household items
208	paper	Household items
209	penny	Household items
210	picture	Household items
211	pillow	Household items
212	plant	Household items
213	plate	Household items
214	purse	Household items
215	radio	Household items
216	rubbish	Household items

217	scissors	Household items
218	soap	Household items
219	spoon	Household items
220	telephone	Household items
221	toothbrush	Household items
222	towel	Household items
223	watch	Household items
224	aunt	People
225	baby	People
226	boy	People
227	brother	People
228	child	People
229	daddy	People
230	doctor	People
231	friend	People
232	person	People
233	policeman	People
234	sister	People
235	girl	People
236	grandma	People
237	grandpa	People
238	lady	People
239	man	People

240	mummy	People
241	nanny	People
242	people	People
243	teacher	People
244	uncle	People

245	bath	Games and Routines
246	breakfast	Games and Routines
247	bye bye	Games and Routines
248	dinner	Games and Routines
249	don’t	Games and Routines
250	hello	Games and Routines
251	hi	Games and Routines
252	lunch	Games and Routines
253	nap	Games and Routines
254	night night	Games and Routines
255	no	Games and Routines
256	pat-a-cake	Games and Routines
257	peekaboo	Games and Routines
258	please	Games and Routines
259	shh / hush / shush	Games and Routines
260	tea	Games and Routines
261	thank you	Games and Routines
262	wait	Games and Routines

263	want to	Games and Routines
264	yes	Games and Routines

265	bite	Action Words
266	blow	Action Words
267	break	Action Words
268	bring	Action Words
269	bump	Action Words
270	call	Action Words
271	carry	Action Words
272	catch	Action Words
273	chase	Action Words
274	clean	Action Words
275	cry	Action Words
276	cuddle	Action Words
277	cut	Action Words
278	dance	Action Words
279	draw	Action Words
280	drink	Action Words
281	drive	Action Words
282	drop	Action Words
283	eat	Action Words
284	fall	Action Words
285	feed	Action Words

286	find	Action Words
287	finish	Action Words
288	get	Action Words
289	give	Action Words
290	go	Action Words
291	have	Action Words
292	hear	Action Words
293	help	Action Words
294	hit	Action Words
295	hug	Action Words
296	hurry	Action Words
297	jump	Action Words
298	kick	Action Words
299	kiss	Action Words
300	know	Action Words
301	like	Action Words
302	look	Action Words
303	love	Action Words
304	make	Action Words
305	open	Action Words
306	play	Action Words
307	pull	Action Words
308	push	Action Words
309	put	Action Words

310	read	Action Words
311	ride	Action Words
312	run	Action Words
313	say	Action Words
314	scratch	Action Words
315	see	Action Words
316	show	Action Words
317	shut / close	Action Words
318	sing	Action Words
319	sleep	Action Words
320	smell	Action Words
321	smile	Action Words
322	splash	Action Words
323	stop	Action Words
324	swim	Action Words
325	swing	Action Words
326	take	Action Words
327	tell	Action Words
328	throw	Action Words
329	tickle	Action Words
330	walk	Action Words
331	wash	Action Words
332	watch	Action Words
333	wipe	Action Words

334	write	Action Words
335	all gone	Descriptive Words
336	asleep	Descriptive Words
337	bad	Descriptive Words
338	big	Descriptive Words
339	blue	Descriptive Words
340	broken	Descriptive Words
341	careful	Descriptive Words
342	clean	Descriptive Words
343	cold	Descriptive Words
344	dark	Descriptive Words
345	dirty	Descriptive Words
346	dry	Descriptive Words
347	empty	Descriptive Words
348	fast	Descriptive Words
349	fine	Descriptive Words
350	gentle	Descriptive Words
351	good	Descriptive Words
352	green	Descriptive Words
353	happy	Descriptive Words
354	hard	Descriptive Words
355	hot	Descriptive Words
356	hungry	Descriptive Words

357	hurt	Descriptive Words
358	little	Descriptive Words
359	nasty	Descriptive Words
360	naughty	Descriptive Words
361	nice	Descriptive Words
362	old	Descriptive Words
363	pretty	Descriptive Words
364	red	Descriptive Words
365	sad	Descriptive Words
366	scared	Descriptive Words
367	sick	Descriptive Words
368	sleepy	Descriptive Words
369	soft	Descriptive Words
370	thirsty	Descriptive Words
371	tired	Descriptive Words
372	wet	Descriptive Words
373	yellow	Descriptive Words

374	how	Question words
375	what	Question words
376	when	Question words
377	where	Question words
378	who	Question words
379	why	Question words

380	day	Time
381	later	Time
382	morning	Time
383	night	Time
384	now	Time
385	today	Time
386	tomorrow	Time
387	tonight	Time

388	her	Pronouns and Possessives
389	his	Pronouns and Possessives
390	I	Pronouns and Possessives
391	it	Pronouns and Possessives
392	me	Pronouns and Possessives
393	mine	Pronouns and Possessives
394	my	Pronouns and Possessives
395	that	Pronouns and Possessives
396	this	Pronouns and Possessives
397	you	Pronouns and Possessives
398	your	Pronouns and Possessives

399	away	Prepositions and Location Words
400	back	Prepositions and Location Words

401	down	Prepositions and Location Words
402	in	Prepositions and Location Words
403	inside	Prepositions and Location Words
404	off	Prepositions and Location Words
405	on	Prepositions and Location Words
406	out	Prepositions and Location Words
407	there	Prepositions and Location Words
408	under	Prepositions and Location Words
409	up	Prepositions and Location Words
410	all	Quantifiers
411	again	Quantifiers
412	another	Quantifiers
413	more	Quantifiers
414	none	Quantifiers
415	not	Quantifiers
416	other	Quantifiers
417	same	Quantifiers
418	some	Quantifiers

Appendix C

Adaptations of Oxford CDI

Note: Only concepts that are common across all language versions of the CDI are listed below. Excludes "Sounds" category.

Table C.1: Oxford CDI items commons across different language versions (category - Animals).

English	Dutch	French	German	Italian	Polish	Portuguese	Spanish
animal	dier	animal	Tier	Animale	zwierze	animal / bicho	animal
bear	beer	ours	Bär	Orso	mis / niedzwiedz	urso	oso
bee	bij	abeille	Biene	Ape	pszczola	abelha	abeja
bird	vogel	oiseau	Vogel	Uccellino	ptaszek	pássaro	pájaro
bunny / rabbit	konijn	lapin	Hase	Coniglio	królik	coelho	conejo / conejito
butterfly	vlinder	papillon	Schmetterling	Farfalla	motyl	borboleta	mariposa
cat	kat	chat	Katze	Gatto	kot	gato	gato
chicken	kuiken	poule	Huhn	Gallina	kurczak	galinha	pollo / pollito
cow	koe	vache	Kuh	Mucca	krowa	vaca	vaca
deer	hert	cerf	Reh	Cervo	jelen	veado	ciervo
dog	hond	chien	Hund	Cane	pies	cão	perro
donkey	ezel	âne	Esel	Asino	osiol	burro	burro
duck	eend	canard	Ente	Papera	kaczka	pato	pato
elephant	olifant	éléphant	Elefant	Elefante	slon	elefante	elefante
fish	vis	poisson	Fisch	Pesce / Pesciolino	ryba	peixe	pez
frog	kikker	grenouille	Frosch	Rana	zaba	sapo	rana

giraffe	giraf	girafe	Giraffe	Giraffa	zyrafa	girafa	jirafa
goose	gans	oie	Gans	Oca	ges	ganso	oca / ganso
horse	paard	cheval	Pferd	Cavallo	kon	cavalo	caballo
kitten	poesje	bébé chat	Kätzchen	Gattino / Micino	kotek	gatinho	gatito
lamb	lam(metje)	agneau	Lamm	Agnello	owieczka	cordeiro	cordero
lion	leeuw	lion	Löwe	Leone	lew	leão	león
monkey	aap	singe	Affe	Scimmia	malpa	macaco	mono
mouse	muis	souris	Maus	Topo	myszka	rato	ratón
owl	uil	hibou	Eule	Gufo	sowa	coruja	búho
penguin	pinguin	pingouin	Pinguin	Pinguino	pingwin	pinguim	pingüino
pig	varken	cochon	Schwein	Maiale	swinka	porco	cerdo / chancho (Arg.) / puerco (Méx.) / cochino / marrano (Col.)
pony	pony	poney	Pony	Pony	kucyk	pónei	poni
puppy	puppy	bébé chien	Hündchen	Cucciolo	szczeniak / piesek	cachorro	perrito
sheep	schaap	mouton	Schaf	Pecora	owca	ovelha	oveja
spider	spin	araignée	Spinne	Ragno	pajak	aranha	araña
squirrel	eekhoorn	écureuil	Eichhörnchen	Scoiattolo	wiewiórka	esquilo	ardilla
tiger	tijger	tigre	Tiger	Tigre	tygrys	tigre	tigre
turkey	kalkoen	dindon	Truthahn	Tacchino	indyk	peru	pavo

turtle	schildpad	tortue	Schildkröte	Tartaruga	zólw	tartaruga	tortuga

Table C.2: Oxford CDI items commons across different language versions (category - Vehicles).

English	Dutch	French	German	Italian	Polish	Portuguese	Spanish
aeroplane / plane	vliegtuig	avion	Flugzeug	Aereo	samolot	avião	avión
bicycle / bike	fiets	vélo	Fahrrad	Bicicletta / Bici	rower	bicicleta	bicicleta / bici
boat	boot	bateau	Boot	Barca	lódz / lódka	barco	barco
bus	bus	bus	Bus	Autobus	autobus	autocarro	autobús / bus (Esp., Col.) / colectivo (Arg.) / guagua (Can.)
car	auto	voiture	Auto	Automobile / Auto / Macchina	auto / samochód	carro / pópó	coche (Esp.) / carro (Méx., Col.) / auto (Arg.)

fire engine	brandweerwagen / brandweerauto	camion de pompier	Feuerwehr(auto)	Autopompa	wóz strazacki	carro de bombeiros	camión de bomberos
lorry / truck	vrachtwagen / camion / vrachtauto	camion	Lastwagen / Laster	Camion	ciezarówka	camião	camión
motor-bike	moto(r) / motorfiets	moto	Motorrad	Motocicletta / Moto	motor / motocykl	moto/a	moto
pushchair / buggy	buggy	poussette	Buggy	Passeggino	wózek	carrinho de bebé / de passeio	cochecito (de bebé)
train	trein	train	Bahn (Eisenbahn, Straßenbahn, U-Bahn, S-Bahn, Zug gelten auch)	Treno	pociag	comboio	tren

Table C.3: Oxford CDI items commons across different language versions (category - Toys).

English	Dutch	French	German	Italian	Polish	Portuguese	Spanish

ball	bal	balle	Ball	Palla	pilka	bola	pelota
balloon	ballon	ballon (de baudruche)	Luftballon	Palloncino	balon(ik)	balão	globo
block / brick	blok	cube	Klotz	Cubi	klocek	cubo	bloque (de juguete)
book	boek	livre	Buch	Libro	ksiazka	livro	libro
bubble	bubbel	bulles	Seifenblase	Bollicina	banka mydlana	bolha (de sabão)	burbuja
doll	pop	poupée	Puppe	Bambola	lalka	boneca	muñeca
pen	pen	stylo	Stift	Penna	pisak / dlugopis	caneta	boli / bolígrafo
teddy bear	teddybeer	nounours	Teddy	Orsacchiotto	mis	urso de peluche	osito de peluche (Col.), osito (Esp.), peluche (Arg.)
toy	speelgoed	jouet	Spielzeug	Giocattolo	zabawka	brinquedo	juguete

Table C.4: Oxford CDI items commons across different language versions (category - Food and Drink).

English	Dutch	French	German	Italian	Polish	Portuguese	Spanish
apple	appel	pomme	Apfel	Mela	jablko	maçã	manzana
banana	banaan	banane	Banane	Banana	banan	banana	plátano (Esp./Méx.) / banano (Col.) / banana (Arg.)

biscuit	beschuit	petits gâteaux	Keks	Biscotti	herbatnik	biscoito	galleta
bread	brood	pain	Brot	Pane	chleb	pão	pan
butter	boter	beurre	Butter	Burro	maslo / margaryna	manteiga	mantequilla
cake	cake	gâteau	Kuchen	Torta	ciasto	bolo	tarta / pastel (Méx.) / torta (Arg.)
carrot	wortel	carotte	Möhre	Carota	marchewka	cenoura	zanahoria
cereal	granen / graansoort	céréales	Müsli	Cereali	platki	cereais	cereales
cheese	kaas	fromage	Käse	Formaggio	ser	queijo	queso
chicken	kip	poulet	Hähnchen	Pollo	kurczak	frango	pollo (comida)
chips	frietjes	frites	Pommes	Patatine	frytki	batatas fritas	patatas fritas
coffee	koffie	café	Kaffee	Caffé	kawa	café	café
drink	drankje	boisson	Getränk	Bibita / Bevanda	picie / napój	bebida	bebida
egg	ei(tje)	oeuf	Eier	Uovo	jajko	ovo	huevo
fish	vis	poisson	Fisch	Pesce	ryba	peixe	pescado
food	eten	nourriture	Essen	Pappa / Cibo	jedzenie / papu / am	papa / comida	comida
ice cream	ijs(je) / crème	glace	Eis	Gelato	lody	gelado	helado

jam	jam	confiture	Marmelade	Marmellata	dzem	geléia / marmelada	mermelada
juice	sap	jus de fruit	Saft	Succo	sok	sumo	zumo
meat	vlees	viande	Fleisch	Carne	mieso	carne	carne (chicha)
milk	melk	lait	Milch	Latte	mleko	leite	leche
orange	appelsien / sinaasappel	orange	Orange	Arancia	pomarancza	laranja	naranja
pasta / spaghetti	pasta / spaghetti	pâtes / spaghetti	Spaghetti	Pasta	makaron / kluski	massa	pasta / espaguetis
peas	erwtjes	petits pois	Erbsen	Piselli	groszek	ervilhas	guisantes
pizza	pizza	pizza	Pizza	Pizza	pizza	piza	pizza
sweets	snoepje	bonbons	Süßigkeiten	Dolci	cukierki	doces	chuches (Esp.) / dulce (Méx.) / golosinas (Arg./Col.)
tea	thee	thé	Tee	Té	herbata	chá	té
toast	geroosterd brood / toast	tartine	Toast	Toast	to(a)st	torrada	tostada
water	water	eau	Wasser	Acqua	woda	água	agua

Table C.5: Oxford CDI items commons across different language versions (category - Body Parts).

English	Dutch	French	German	Italian	Polish	Portuguese	Spanish
belly button / tummy button	navel	nombril	Bauchnabel	Ombelico	pepek	umbigo	ombligo
cheek	wang	joue	Backe	Guancia	policzek	bochecha	mejilla
ear	oor	oreille	Ohr	Orecchio	ucho	ouvido(s)	oreja
eye	oog	yeux	Auge	Occhio	oko	olho(s)	ojo
face	gezicht	figure	Gesicht	Faccia	twarz	cara	cara
finger	vinger	doigt	Finger	Dito	palec	dedo	dedo
foot	voet	pied	Fuß	Piede	stopa	pé(s)	pie
hair	haar	cheveux	Haare	Capelli	wlosy	cabelo	pelo
hand	hand	main	Hand	Mano	reka	mão(s)	mano
head	hoofd	tête	Kopf	Testa	glowa	cabeça	cabeza
knee	knie	genou	Knie	Ginocchio	kolano	joelho	rodilla
leg	been	jambe	Bein	Gamba	noga	perna(s)	pierna
mouth	mond	bouche	Mund	Bocca	usta	boca	boca
nose	neus	nez	Nase	Naso	nos	nariz	nariz
toe	teen	doigt de pied	Zeh	Dito del piede	palec u nogi	dedo do pé	dedo del pie
tongue	tong	langue	Zunge	Lingua	jezyk	língua	lengua

tooth	tand	dent	Zahn	Dente	zab	dente(s)	diente
tummy	buik	ventre	Bauch	Pancia	brzuch	barriga	barriga

Table C.6: Oxford CDI items commons across different language versions (category - Clothes).

English	Dutch	French	German	Italian	Polish	Portuguese	Spanish
bib	slabbetje	bavoir/bavette	Lätzchen	Bavaglino	sliniaczek	babete	babero
boot(s)	laarsjes	bottes	Stiefel	Stivali	kozaki	bota(s)	bota/s
button	knoop	boutons	Knopf	Bottone	guzik	botão	botón
coat	jas / overjas	manteau	Mantel	Cappotto	plaszcz	casaco	abrigo
dress	kleedje / jurkje	robe	Kleid	Vestito	sukienka	vestido	vestido
glasses / specs	bril	lunettes	Brille	Occhiali	okulary	óculos	gafas
hat	hoed	chapeau	Mütze	Cappello	kapelusz	chapéu	sombrero / gorro/a
jacket	jasje	veste	Jacke	Giacca	kurtka	kispo / blusão	chaqueta
jeans	jeans / spijkerbroek	jeans	Jeans	Jeans	dzinsy	jeans	vaquero / tejano
jumper / sweater	trui / sweater / pull	pull / sweat	Pullover	Maglione	sweter	camisola	jersey / suéter

nappy	luier	couche	Windel	Pannolino	pieluszka / pampers / pielucha	fralda	pañal
necklace	(hals)ketting	collier	Kette	Collana	naszyjnik / wisiorek	colar / fio	collar
pyjamas	pyjama	pyjama	Schlafanzug	Pigiama	pizama	pijama	pijama
shirt	hemd	chemise	Hemd	Camicia	koszula	camisa	camisa
shoe	schoen	chaussure	Schuh	Scarpa	buty	sapato(s)	zapato/s
shorts	short / korte broek	short	Shorts	Pantaloncini	szorty	calções	pantalones cortos / shorts / bermudas / guayucos (Col.)
sock	sok	chaussettes	Socken	Calze / Calzini	skarpetki	meia(s)	calcetín
trousers	(lange) broek	pantalon	Hose	Pantaloni	spodnie	calças	pantalón
zip	rits	fermeture éclair	Reißverschluss	Cerniera	zamek blyskawiczny	fecho	cremallera

Table C.7: Oxford CDI items commons across different language versions (category - Furniture and Rooms).

English	Dutch	French	German	Italian	Polish	Portuguese	Spanish
bath / bathtub	bad	baignoire	Badewanne	Vasca da bagno	wanna	banheira	bañera
bathroom	badkamer	salle de bain	Badezimmer / Bad	Bagno	lazienka	casa / quarto de banho	baño
bed	bed	lit	Bett	Letto	lózko	cama	cama
bedroom	slaapkamer	chambre	Schlafzimmer	Camera	sypialnia	quarto	habitación / cuarto (de dormir)
chair	stoel	chaise	Stuhl	Sedia	krzeslo	cadeira	silla
cooker / stove / oven	fornuis / vuur / kookplaat	cuisinière	Herd	Cucina / Fornelli / Forno	kuchenka / piekarnik	fogão / forno	horno / estufa
cot	wieg	berceau	Kinderbettchen	Culla	lózeczko	berço	cuna
door	deur	porte	Tür	Porta	drzwi	porta	puerta
drawer	lade / schuif	tiroir	Schublade	Cassetto	szuflada	gaveta	cajón
garage	garage	garage	Garage	Garage	garaz	garagem	garaje
high chair	kinderstoel	chaise haute	Hochstuhl	Seggiolone	wysokie krzeslo	cadeira alta / da papa	trona (Esp.) / silla de bebés (Col.)

kitchen	keuken	cuisine	Küche	Cucina	kuchnia	cozinha	cocina
living room	woonkamer / living / huiskamer	salon	Wohnzimmer	Salotto	salon / duzy pokój	sala (de estar)	salón
play pen	(baby)box	parc	Laufstall	Box	kojec (dla dziecka)	parque (em casa)	parque (en casa) corralito
potty	potje	pot	Töpfchen	Vasino	nocnik	bacio / penico	orinal (Es.) / bacinica (Méx.) / bacinilla (Col.) / pelela (Arg.)
refrigerator / fridge	koelkast / frigo / ijskast	frigo	Kühlschrank	Frigorifero / Frigo	lodówka	frigorífico	nevera / frigo
rocking chair	schommelstoel	rocking chair / chaise à bascule	Schaukelstuhl	Sedia a dondolo	bujany fotel / fotel na biegunach / bujak	cadeira de balanço	mecedora (Esp.) / silla mecedora (Col.)
settee / sofa	divan / sofa / bank	canapé	Sofa	Divano	fotel	sofá	sofá (Esp., Méx., Col.) / sillón (Arg.)

sink	gootsteen / afwasbak	évier	Waschbecken	Lavandino	umywalka	lava-loiça / lavatório	lavabo
stairs	trap	escalier	Treppe	Scale	schody	escadas	escalera/s
table	tafel	table	Tisch	Tavolo	stól	mesa	mesa
TV / television	TV / televisie	télé	Fernseher	Televisione / TV	telewizor	televisão / TV	televisión / tele
window	raam / venster	fenêtre	Fenster	Finestra	okno	janela	ventana

Table C.8: Oxford CDI items commons across different language versions (category - Outside).

English	Dutch	French	German	Italian	Polish	Portuguese	Spanish
beach	strand	plage	Strand	Spiaggia	plaza	praia	playa
bucket	emmer	seau	Eimer	Secchiello	wiaderko	balde	cubo
church	kerk	église	Kirche	Chiesa	kosciól	igreja	iglesia
flower	bloem	fleur	Blume	Fiore	kwiatek	flôr	flor
garden	tuin	jardin	Garten	Giardino	ogród	jardim	jardín
house	huis	maison	Haus	Casa	dom	casa	casa
moon	maan	lune	Mond	Luna	ksiezyc	lua	luna

party	feestje	fête	Party	Festa	przyjecie / impreza	festa	fiesta
pool	zwembad	piscine	Schwimmbad	Piscina	basen	piscina	piscina
rain	regen	pluie	Regen	Pioggia	deszcz	chuva	lluvia
school	school	école	Schule	Scuola	szkola	escola / colégio	cole (Esp.) / colegio (Arg.) / escuela (Mex., Col.)
shop	winkel	magasin	Geschäft	Negozio	sklep	loja	tienda (Esp., Méx., Col.) / negocio (Arg.)
sky	lucht / hemel	ciel	Himmel	Cielo	niebo	céu	cielo
slide	glijbaan	toboggan	Rutsche	Scivolo	slizgawka	escorrega	tobogán
snow	sneeuw	neige	Schnee	Neve	snieg	neve	nieve
spade	schep	pelle	Schaufel	Paletta	lopat(k)a	pá	pala
star	ster	étoile	Stern	Stella	gwiazda	estrela	estrella
stone	steen	caillou	Stein	Sasso	kamien	pedra	piedra
sun	zon	soleil	Sonne	Sole	slonce	sol	sol
swing	schommel	balançoire	Schaukel	Altalena	hustawka	baloiço	columpio (Esp., Méx., Col.) / hamaca (Arg.)
tree	boom	arbre	Baum	Albero	drzewo	árvore	árbol
wall	muur / wand	mur	Mauer / Wand	Muro	sciana	parede	pared

water	water	eau	Wasser	Acqua	woda	água	agua
work	werk	travail	Arbeit	Lavoro	praca	trabalho	trabajo
zoo	zoo / dierentuin	zoo	Zoo	Zoo	ogród zoologiczny / zoo	jardim zoológico	zoo

Table C.9: Oxford CDI items commons across different language versions (category - Household Items).

English	Dutch	French	German	Italian	Polish	Portuguese	Spanish
bin	vuilnisbak	poubelle	Mülleimer	Cestino	kosz	caixote do lixo	papelera
blanket	deken	couverture	Decke	Coperta	koc / koldra	manta	manta (Esp.) / colcha (Arg.) / cobija (Col.)
bottle	fles	bouteille	Flasche	Bottiglia	butelka	garrafa	botella
bowl	kom / schaal	bol	Schüssel	Scodella	miska	tigela	cuenco / bol
box	doos	boîte	Kiste	Scatola	pudelko	caixa	caja
broom	bezem	balai	Besen	Scopa	miotla	vassoura	escoba
brush	borstel	brosse	Bürste	Pennello	szczotka	escova	cepillo
clock	klok	horloge	Uhr	Orologio	zegar	relógio	reloj
comb	kam	peigne	Kamm	Pettine	grzebien	pente	peine

cup	kop	tasse	Becher	Tazza	filizanka	chávena	taza
dummy	speen	sucette	Schnuller	Ciuccio	smoczek	chupeta / chucha	chupete
fork	vork	fourchette	Gabel	Forchetta	widelec	garfo	tenedor
glass	glas	verre	Glas	Bicchiere	szklanka	copo	vaso
hammer	hamer	marteau	Hammer	Martello	mlotek	martelo	martillo
hoover / vacuum cleaner	stofzuiger	aspirateur	Staubsauger	Aspirapolvere	odkurzacz	aspirador	aspiradora
jug	kruik	cruche / pot à eau / pichet	Kanne	Caraffa	dzbanek	jarro	jarra
key	sleutel	clefs	Schlüssel	Chiave	klucz	chave	llaves
lamp	lamp	lampe	Lampe	Lampada	lampa	candeeiro	lámpara
light	licht	lumière	Licht	Luce	swiatlo	luz	luz
medicine	medicijn(en) (siroop, pilletje, drankje)	médicaments	Medizin	Medicina	lekarstwa / tabletki	medicamento / remédio	medicina
money	geld	argent	Geld	Soldi	pieniadze	dinheiro	dinero
paper	papier	papier	Papier	Carta	kartka / papier	papel	papel
picture	foto	photo	Bild	Fotografia / Foto	zdjecie / fotografia	fotografia	foto

pillow	hoofdkussen	oreiller	Kissen	Cuscino	poduszka	almofada	cojín / almohada
plant	plant	plante	Pflanze	Pianta	roslina	planta	planta
plate	bord	assiette	Teller	Piatto	talerz	prato	plato
purse	portemonnee	porte-monnaie	Geldbeutel	Portafoglio	portfel	carteira	bolso
radio	radio	radio	Radio	Radio	radio	rádio	radio
scissors	schaar	ciseaux	Schere	Forbici	nozyczki	tesoura	tijeras
soap	zeep	savon	Seife	Sapone	mydlo	sabão / sabonete	jabón
spoon	lepel	cuillère	Löffel	Cucchiaio	lyzka	colher	cuchara
telephone	telefoon	téléphone	Telefon	Telefono	telefon	telefone	teléfono
toothbrush	tandenborstel	brosse à dent	Zahnbürste	Spazzolino da denti	szczoteczka do zebów	escova de dentes	cepillo de dientes
towel	handdoek	serviette	Handtuch	Asciugamano	recznik	toalha	toalla

Table C.10: Oxford CDI items commons across different language versions (category - People).

English	Dutch	French	German	Italian	Polish	Portuguese	Spanish
aunt	tante	tante	Tante	Zia	ciocia	tia	tía
baby	baby	bébé	Baby	Bimbo/a	dzidzia / dzidzius / dziecko	bebé	bebé

boy	jongen	garçon	Junge	Ragazzo	chlopiec	menino	chico / nene
brother	broer	frère	Bruder	Fratello	brat	irmão	hermano
child	kind	enfant	Kind	Bambino/a	dziecko	criança	niño
daddy	papa	papa	Papa	Papà	tata / tatus	papá	papá (papi)
doctor	dokter	docteur	Doktor / Arzt	Dottore	doktor / lekarz	médico / doutor	doctor/a
friend	vriend(in)	ami/e, copain / copine	Freund / Freundin	Amico/a	przyjaciel	amigo	amigo/a
girl	meisje	fille	Mädchen	Ragazza	dziewczynka	menina	chica / nena
grandma	oma	grand-mère	Oma	Nonna	babcia	avó	abuela (yaya)
grandpa	opa	grand-père	Opa	Nonno	dziadek / dziadzio	avô	abuelo (yayo)
lady	mevrouw	dame	Frau	Signora	pani	senhora	señora
man	meneer	monsieur	Mann	Signore	pan	senhor	señor
mummy	mama	maman	Mama	Mamma	mama / mamusia	mamã	mamá (mami)
nanny	babysitter	nounou	Kindermädchen	Babysitter / Tata	niania / opiekunka	ama	niñera / canguro
people	mensen	gens	Menschen	Gente	ludzie	gente	gente
person	persoon	personne	Person	Persona	osoba	pessoa	persona
policeman	politieagent	police	Polizei	Poliziotto	policjant	polícia	policía
sister	zus	soeur	Schwester	Sorella	siostra	irmã	hermana

teacher	schooljuffrouw / lerares / juf	maître / sse	Lehrer	Maestro/a	nauczyciel	professor / a	maestra/o
uncle	oom / nonkel	oncle	Onkel	Zio	wujek	tio	tío

Table C.11: Oxford CDI items commons across different language versions (category - Games and Routines).

English	Dutch	French	German	Italian	Polish	Portuguese	Spanish
bath	een bad nemen	bain	baden	Fare il bagno	kapiel	para o banho!	(vamos a) a bañar
breakfast	ontbijt	petit déjeuner	Frühstück	Colazione	sniadanie	pequeno-almoço	(vamos a) desayunar
dinner	diner	dîner	Abendbrot	Cena	kolacja	jantar / ceia	(vamos a) cenar
don't	niet doen	ne fais pas	(tu das) nicht	Non fare	nie rób tego	não! (isso não)	¡no! (eso no)
hello	hallo	bonjour	hallo	Buongiorno	halo	olá!	hola
lunch	lunch	déjeuner	Mittagessen	Pranzo	obiad	almoço	(vamos a) comer / almorzar
nap	slaapje doen	sieste	Mittagsschlaf	Pisolino	drzemka	soneca / sesta	siesta
night night	slaap lekker / welterusten	bonne nuit	Gute Nacht	Buonanotte	dobranoc	boa noite	buenas noches
no	nee / neen	non	nein	No	nie	não	no

peekaboo	kiekeboe	coucou	guck guck	Cucù/Bubù-settete	a kuku	teté / jogar às escondidas	cu-cu (juego)
please	alsjeblieft	s'il te plaît	bitte	Prego / Per favore	prosze	por favor!	por favor
shh / hush / shush	sst	chut	Sschhhh / Psst	Zitto / Shhh	ciii / sza	chiu / shhh	shh shh (silencio)
tea	avondeten	goûter	Kafe und Kuchen	Merenda	przekaska / podwieczorek	lanche	merienda
thank you	dank u / dank u wel	merci	danke	Grazie	dziekuje	obrigado / a!	gracias
wait	wacht!	attends	warten	Aspettare	czekac	espera!	espera
want to	willen	veux	möchte	Volere	chciec	querer	querer (desear)
yes	ja	oui	ja	Sì	tak	sim	sí (afirmación)

Table C.12: Oxford CDI items commons across different language versions (category - Action Words).

English	Dutch	French	German	Italian	Polish	Portuguese	Spanish
bite	bijten	mordre	beißen	Mordere	ugryzc	morder	morder
blow	blazen	souffler	pusten	Soffiare	dmuchac	soprar	soplar

break	breken	casser	kaputt machen	Rompere	zlamac	partir / estragar	romper(se)
bring	brengen	apporter	bringen	Portare	przyniesc	trazer	traer
call	bellen	appeler	anrufen	Chiamare	dzwonic	chamar	llamar
catch	vangen	attraper	fangen	Acchiappare	zlapac	apanhar	coger / agarrar
chase	achtervolgen	chasser	jagen	Inseguire	gonic	perseguir	perseguir
cry	huilen	pleurer	weinen	Piangere	plakac	chorar	llorar
cuddle	knuffelen	caliner (faire un calin)	kuscheln	Coccolare	tulic	acariciar / dar mimo	hacer/dar un mimo/mimitos
dance	dansen	danser	tanzen	Ballare	tanczyc	dançar	bailar
drink	drinken	boire	trinken	Bere	pic	beber	beber
drive	rijden	conduire	fahren	Guidare	prowadzic	conduzir	conducir
drop	druppel	laisser tomber	fallen lassen	Lasciare / Fare cadere	upuscic	soltar / deixar cair	dejar caer
eat	eten	manger	essen	Mangiare	jesc	comer	comer
fall	vallen	tomber	fallen	Cadere	upasc / spasc	cair	caer(se)
find	vinden	trouver	finden	Trovare	znalezc	encontrar	encontrar
finish	afhebben	finir	fertig machen	Finire	skonczyc	acabar	acabar / terminar
get	krijgen	recevoir	bekommen	Prendere	dostac	pegar	guardar
give	geven	donner	geben	Dare	dac	dar	dar

go	gaan	aller	laufen	Andare	pójsc	ir	ir
have	hebben	avoir	haben	Avere	miec	ter	tener
hear	horen	entendre	hören	Sentire	slyszec	escutar / ouvir	oír
help	helpen	aider	helfen	Aiutare	pomóc	ajudar	ayudar
hit	slaan	taper	hauen	Colpire	uderzyc	bater	pegar (un golpe)
hurry	zich haasten / zich spoeden	se dépêcher	beeilen	Sbrigarsi	spieszyc sie	apressar	darse prisa
jump	springen	sauter	springen	Saltare	skakac	saltar	saltar
kick	schoppen	donner un coup (de pied)	treten	Dare un calcio	kopac	chutar	dar una patada / patear
kiss	kusje geven / zoenen	faire un bisou	küssen	Baciare	pocalowac	beijar	besar
know	kennen	savoir	wissen	Conoscere / Sapere	wiedziec	saber	saber
like	leuk vinden	aimer bien	lieb haben	Piacere	lubic	gostar	gustar (me gusta)
love	houden van	aimer	lieben	Amare	kochac	amar	querer
make	maken	faire	machen	Fare	robic	fazer	hacer
open	openen	ouvrir	aufmachen	Aprire	otworzyc	abrir	abrir
play	spelen	jouer	spielen	Giocare	grac	brincar	jugar
pull	trekken	tirer	ziehen	Tirare	ciagnac	puxar	tirar (hacia uno)

push	duwen	pousser	drücken	Spingere	popychac	empurrar	empujar
put	zetten	mettre	(hin) stellen	Mettere	polozyc	pôr	poner(se)
read	lezen	lire	lesen	Leggere	czytac	ler	leer
run	rennen	courir	rennen	Correre	biec / biegac	correr	correr
say	zeggen	dire	sagen	Dire	mówic	dizer	decir
scratch	krabben	gratter	kratzen	Grattare / Graffiare	(za)drapac	coçar	arañar
see	zien	voir	sehen	Vedere	widziec	ver	ver
show	tonen / wijzen / laten zien	montrer	zeigen	Mostrare	pokazywac	mostrar	mostrar / enseñar (un objeto)
shut / close	sluiten	fermer	zumachen	Chiudere	zamknac	fechar	cerrar
sing	zingen	chanter	singen	Cantare	spiewac	cantar	cantar
sleep	slapen	dormir	schlafen	Dormire	spac	dormir	dormir(se)
smell	ruiken	sentir	riechen	Annusare	smierdziec	cheirar	oler
smile	glimlachen	sourire	lächeln	Sorridere	usmiechac sie	sorrir	sonreír
splash	spatten / spetteren	éclabousser	spritzen	Schizzare / Spruzzare	prysnac / opryskac / pryskac	salpicar	salpicar
stop	toppen / ophouden	arrêter	aufhören	Fermare	zatrzymac	parar	parar
swim	zwemmen	nager	schwimmen	Nuotare	plywac	nadar	nadar

swing	schommelen	balancer	schaukeln	Dondolare	hustac sie / bujac	baloiçar	columpiar(se)
throw	gooien / werpen	jeter	werfen	Buttare	rzucac	atirar	tirar
tickle	kietelen	chatouiller	kitzeln	Solleticare	laskotac	fazer cócegas	(hacer / tener) cosquillas
walk	wandelen	marcher	gehen	Camminare	isc	andar	caminar
wash	wassen	laver	waschen	Lavare	myc (sie)	lavar	lavar(se)
write	schrijven	écrire	schreiben	Scrivere	pisac	escrever	escribir

Table C.13: Oxford CDI items commons across different language versions (category - Descriptive Words).

English	Dutch	French	German	Italian	Polish	Portuguese	Spanish
all gone	alles weg / op	fini! / c'est fini! / c'est parti!	alle (im Sinne von ”weg”)	Tutto finito	nie ma nic	acabou / já nao há	no está (en un juego)
asleep	in slaap	endormi	schlafend	Addormentato	spiacy	a dormir	dormido

bad	slecht	pas bon / pas bien / mal	schlecht	Cattivo	zly	mau / má	malo
big	groot	grand	groß	Grande / Grosso	duzy	grande	grande
blue	blauw	bleu	blau	Blu	niebieski	azul	azul
broken	kapot / stuk	cassé	kaputt	Rotto	zlamany	estragado / partido	roto
careful	voorzichtig	attention	vorsichtig	Prudente	ostrozny	cuidadoso / atento	cuidadosa/o
cold	koud	froid	kalt	Freddo	zimny / zimno	frio	frío
dark	donker	sombre	dunkel	Buio	ciemno	escuro	oscuro
dirty	vuil	sale	schmutzig	Sporco	brudny	sujo	sucio
dry	droog	sec / che	trocken	Asciutto	suchy	seco	seco
empty	leeg	vide	leer	Vuoto	pusty	vazio	vacío
fast	snel / vlug	vite	schnell	Veloce / Rapido	predko / szybko	depressa / rápido	rápido
fine	fijn	bien	fein	Bene	w porzadku	bem	está bien
gentle	aardig	tendre	sanft	Gentile	delikatny	delicado	amable
good	goed	bon / ne	gut	Buono	dobry / dobrze	bom	bueno
green	groen	vert	grün	Verde	zielony	verde	verde
happy	blij	content / te	froh	Felice	szczesliwy	contente / feliz	contento / feliz
hard	hard	dur	hart	Duro	ciezko	duro	duro
hot	heet	chaud / e	heiß	Caldo	cieply / cieplo	quente	caliente
how	hoe	comment	wie	Come?	jak?	como?	cómo

hungry	honger (hebben)	avoir faim	hungrig	Affamato	glodny	com fome	hambre
hurt	gewond	blessé	verletzt	Ferito	to boli	magoado	hacerse daño
little	klein	petit / e	klein	Piccolo	maly	pequeno	pequeño / chiquito
naughty	stout	vilain / e	ungezogen	Disubbidiente	niegrzeczny	mal comportado	travieso
old	oud	vieux / vielle	alt	Vecchio	stary	velho	viejo
red	rood	rouge	rot	Rosso	czerwony	vermelho	rojo
sad	bedroefd	triste	traurig	Triste	smutny	triste	triste
scared	bang	avoir peur	ängstlich	Spaventato	przerazony / wystraszony	assustado	asustado
sick	ziek	malade	krank	Malato	chory	doente / mal disposto	enfermo / malito
sleepy	slaperig	avoir sommeil	schläfrig	Assonnato	senny	ensonado / sonolento	tener sueño
soft	zacht	doux/ce	weich	Morbido	miekki	suave	suave / blando
thirsty	dorst (hebben)	avoir soif	durstig	Assetato	spragniony / chce mi sie pic	com sede / sedento	tener sed
tired	moe	fatigué	müde	Stanco	zmeczony	cansado	cansado
wet	nat	mouillé	nass	Bagnato	mokry / mokro	molhado	mojado
yellow	geel	jaune	gelb	Giallo	zólty	amarelo	amarillo

Table C.14: Oxford CDI items commons across different language versions (category - Question Words).

English	Dutch	French	German	Italian	Polish	Portuguese	Spanish
what	wat	quoi	was	Che / Che cosa?	co?	o quê?	qué
when	wanneer	quand	wann	Quando?	kiedy?	quando?	cuándo
where	waar	où	wo	Dove?	gdzie?	onde?	dónde
who	wie	qui	wer	Chi?	kto?	quem?	quién
why	waarom	pourquoi	warum	Perchè?	dlaczego? / czemu?	porquê?	por qué

Table C.15: Oxford CDI items commons across different language versions (category - Time).

English	Dutch	French	German	Italian	Polish	Portuguese	Spanish
day	dag	jour	Tag	Giorno	dzien	dia	día
later	later / straks	plus tard / tout à l'heure	später	(più) Tardi / Dopo	potem / pózniej	mais tarde	después
morning	s morgens / ochtend	matin	Morgen	Mattina	ranek / rano	manhã	mañana
night	nacht	nuit	Nacht	Notte	noc / w nocy	noite	noche
now	nu	maintenant	jetzt	Adesso / Ora	teraz	agora	ahora
today	vandaag	aujord'hui	heute	Oggi	dzisiaj	hoje	hoy

tonight	vannacht	ce soir	heute Abend	Stasera / Stanotte	dzisiaj wieczorem	hoje de noite	esta noche

Table C.16: Oxford CDI items commons across different language versions (category - Pronouns).

English	Dutch	French	German	Italian	Polish	Portuguese	Spanish
I	ik	je	ich	Io	ja	eu	yo
me	mij	moi	mich	Me	mnie, mi	mim	mí (para mí)
mine	van mij	à moi	meins	Mio/a	mój	meu / minha	mío / mía
you	jij / je	vous/tu	du	Tu	ty / ci	tu	tu
your	jouw(e) / van jou / de jou(we)	votre / ta / ton	dein	Tuo/a	twój	tua / teu	tu (tu casa)

Table C.17: Oxford CDI items commons across different language versions (category - Prepositions and Location Words).

English	Dutch	French	German	Italian	Polish	Portuguese	Spanish
away	weg	loin	weg	Lontano	daleko	ausente	lejos
back	terug	derrière	zurück	Dietro	z powrotem	detrás / atrás	atrás / detrás
down	beneden	en bas	runter	Giù	na dół / na dole	abaixo	abajo
in	in	dans	im	In	w	em / na / no	en

inside	binnen(in)	dedans	drinnen	Dentro	wewnatrz	dentro	dentro / adentro
there	daar	là-bas	da	Lì / Là	tam	ali / lá	ahí / allí
under	onder	sous	unter	Sotto	pod	abaixo de / sob	debajo
up	omhoog	en haut	oben	Sopra	na górze	em cima	arriba

Table C.18: Oxford CDI items commons across different language versions (category - Quantifiers).

English	Dutch	French	German	Italian	Polish	Portuguese	Spanish
again	opnieuw	encore	wieder	Ancora	znowu	outra vez / de novo	otra vez
all	alles / allemaal	tous / tout	alle	Tutto	wszystko / wszyscy	todo(a) / todos(as) / tudo	todo / todas/os
more	meer	plus	mehr	Più / Di più	wiecej	mais	más
none	geen / niemand / niets	aucun / ne	kein	Nessuno	zaden	ninguém / nenhum	ninguno
other	ander	autre	andere	Altro	inny	outro	otro
same	hetzelfde	le / la même	dasselbe	Stesso	ten sam	mesmo	mismo / igual
some	een beetje	un peu	manche	Qualche	jakis	algum	algunos/as

Appendix D

Translation Elicitation Task

D.1 Included trials

D.1.1 Spanish-English

Table D.1: List of presented words (Spanish), with translation in English, percentage of English participants who gave correct answer, Levenshtein similarity scores and ALINE similarity score.

word	translation	%correct	lv	aline
guitarra	guitar	100.0	0.333	0.868
tomate	tomato	100.0	0.375	0.848
chocolate	chocolate	96.6	0.500	0.831

sandwich	sandwich	96.6	0.571	0.818
patata	potato	89.7	0.375	0.861
pingüino	penguin	79.3	0.625	0.945
tigre	tiger	56.7	0.200	0.696
naranja	orange	44.8	0.000	0.631
tren	train	43.3	0.600	0.941
chaqueta	jacket	40.0	0.333	0.842
plato	plate	40.0	0.400	0.867
boton	button	36.7	0.600	0.878
ensalada	salad	34.5	0.375	0.776
leon	lion	31.0	0.400	0.884
cebra	zebra	24.1	0.167	0.891
pollo	chicken	24.1	0.000	0.179
botella	bottle	20.7	0.333	0.770
flor	flower	20.7	0.400	0.723
bol	bowl	16.7	0.500	0.919
gato	cat	16.7	0.250	0.778
perro	dog	13.8	0.000	0.314
jirafa	giraffe	13.3	0.000	0.793
mesa	table	10.3	0.200	0.315
pluma	feather	10.3	0.000	0.417
puerta	door	10.3	0.000	0.288
mariposa	butterfly	10.0	0.000	0.600
bici	bike	6.9	0.250	0.650

esponja	sponge	6.9	0.286	0.712
frambuesa	raspberry	6.9	0.111	0.538
gafas	glasses	6.9	0.286	0.706
pie	foot	6.9	0.000	0.387
moto	motorbike	6.7	0.222	0.584
bocadillo	sandwich	3.4	0.000	0.457
caballo	horse	3.4	0.000	0.232
galleta	biscuit	3.4	0.000	0.463
globo	balloon	3.4	0.167	0.582
mano	hand	3.4	0.250	0.491
pato	duck	3.4	0.000	0.605
camiseta	t-shirt	3.3	0.125	0.729
manzana	apple	3.3	0.000	0.237
pelota	ball	3.3	0.000	0.520
platano	banana	3.3	0.143	0.629
avestruz	ostrich	0.0	0.250	0.800
barco	boat	0.0	0.200	0.613
biberon	bottle	0.0	0.143	0.420
boca	mouth	0.0	0.000	0.520
boli	pen	0.0	0.000	0.400
brazo	arm	0.0	0.000	0.250
buho	owl	0.0	0.000	0.231
burro	donkey	0.0	0.000	0.289
caja	box	0.0	0.000	0.545

calcetin	sock	0.0	0.000	0.459
camisa	shirt	0.0	0.000	0.528
caracol	snail	0.0	0.143	0.419
cepillo	brush	0.0	0.000	0.319
cerdo	pig	0.0	0.000	0.400
cereza	cherry	0.0	0.000	0.744
cesta	basket	0.0	0.143	0.588
cielo	sky	0.0	0.000	0.484
coche	car	0.0	0.250	0.606
cubo	bucket	0.0	0.000	0.417
cuchara	spoon	0.0	0.000	0.253
dedo	finger	0.0	0.000	0.374
estrella	star	0.0	0.286	0.667
gallina	cock	0.0	0.000	0.357
gorrion	sparrow	0.0	0.000	0.427
gorro	hat	0.0	0.000	0.162
grifo	tap	0.0	0.000	0.436
gusano	worm	0.0	0.000	0.513
hormiga	ant	0.0	0.000	0.306
ladrillo	brick	0.0	0.000	0.518
lapiz	pencil	0.0	0.000	0.566
loro	parrot	0.0	0.000	0.383
martillo	hammer	0.0	0.000	0.330
moneda	coin	0.0	0.000	0.400

mono	monkey	0.0	0.200	0.596
muneca	doll	0.0	0.000	0.289
nariz	nose	0.0	0.200	0.655
ombligo	belly button	0.0	0.125	0.499
osito	teddy bear	0.0	0.143	0.500
oveja	sheep	0.0	0.000	0.409
pajaro	bird	0.0	0.000	0.424
panal	nappy	0.0	0.000	0.357
pastel	cake	0.0	0.000	0.415
pez	fish	0.0	0.000	0.776
pierna	leg	0.0	0.000	0.255
tambor	drum	0.0	0.000	0.497
tenedor	fork	0.0	0.000	0.288
tijeras	scissors	0.0	0.000	0.575
tobogan	slide	0.0	0.000	0.463
ventana	window	0.0	0.000	0.489
zanahoria	carrot	0.0	0.000	0.476
arbol	tree	0.0	0.200	0.315
ola	wave	0.0	0.000	0.276

D.1.2 Catalan-English

Table D.2: List of presented words (Catalan), with translation in English, percentage of English participants who gave correct answer, Levenshtein similarity scores and ALINE similarity score.

word	translation	%correct	lv	aline
patata	potato	100.0	0.625	0.891
guitarra	guitar	97.1	0.333	0.845
xocolata	chocolate	97.1	0.375	0.805
zebra	zebra	91.2	0.333	0.912
girafa	giraffe	85.3	0.000	0.900
pingui	penguin	85.3	0.571	0.878
planta	plant	82.4	0.500	0.912
esponja	sponge	79.4	0.429	0.865
tomaquet	tomato	73.5	0.375	0.751
jaqueta	jacket	55.9	0.333	0.891
tigre	tiger	55.9	0.400	0.800
mirall	mirror	44.1	0.200	0.826
tren	train	35.3	0.400	0.941
flor	flower	26.5	0.400	0.850
forquilla	fork	26.5	0.286	0.632
bol	bowl	20.6	0.500	0.919
gat	cat	20.6	0.333	0.847
cavall	horse	14.7	0.000	0.289
boto	button	11.8	0.500	0.682

camisa	shirt	11.8	0.000	0.528
poma	apple	11.8	0.000	0.378
bici	bike	8.8	0.250	0.650
cama	bed	8.8	0.000	0.411
orella	ear	8.8	0.200	0.372
papallona	butterfly	8.8	0.125	0.650
plat	plate	8.8	0.600	0.918
porta	door	8.8	0.200	0.410
boca	mouth	5.9	0.000	0.520
bou	bull	5.9	0.333	0.667
cotxe	car	5.9	0.250	0.667
moto	motorbike	5.9	0.222	0.571
nas	nose	5.9	0.250	0.778
pollastre	chicken	5.9	0.000	0.372
taronja	orange	5.9	0.000	0.713
arbre	tree	2.9	0.000	0.590
bibero	bottle	2.9	0.167	0.474
estrella	star	2.9	0.286	0.667
estruc	ostrich	2.9	0.333	0.900
formiga	ant	2.9	0.000	0.326
fulla	leaf	2.9	0.000	0.440
galleda	bucket	2.9	0.000	0.414
mico	monkey	2.9	0.400	0.723
moneda	coin	2.9	0.000	0.376

pa	bread	2.9	0.000	0.412
ungla	nail	2.9	0.200	0.511
abella	bee	0.0	0.000	0.522
aixeta	tap	0.0	0.000	0.620
amanida	salad	0.0	0.000	0.333
ampolla	bottle	0.0	0.000	0.615
berenar	snack	0.0	0.000	0.415
bolet	mushroom	0.0	0.000	0.380
bolquer	nappy	0.0	0.000	0.426
caixa	box	0.0	0.000	0.600
cirera	cherry	0.0	0.000	0.728
colom	pigeon	0.0	0.000	0.489
cuc	worm	0.0	0.000	0.121
dit	finger	0.0	0.000	0.373
entrepa	sandwich	0.0	0.000	0.370
galeta	biscuit	0.0	0.000	0.506
gerd	grass	0.0	0.000	0.620
globus	balloon	0.0	0.167	0.538
gorra	cap	0.0	0.000	0.378
got	cup	0.0	0.000	0.718
lleo	lion	0.0	0.200	0.600
lloro	parrot	0.0	0.000	0.357
ma	hand	0.0	0.000	0.341
mussol	owl	0.0	0.200	0.500

ovella	sheep	0.0	0.000	0.279
pernil	ham	0.0	0.000	0.306
peu	foot	0.0	0.000	0.440
ploma	feather	0.0	0.200	0.460
raspall	brush	0.0	0.000	0.559
samarreta	t-shirt	0.0	0.125	0.564
tisores	scissors	0.0	0.286	0.719
tobogan	slide	0.0	0.000	0.463
ull	eye	0.0	0.000	0.125
ulleres	glasses	0.0	0.000	0.513
vaixell	boat	0.0	0.400	0.604
onada	wave	0.0	0.000	0.564

D.1.3 Catalan-Spanish

Table D.3: List of presented words (Catalan), with translation in Spanish, percentage of Spanish participants who gave correct answer, Levenshtein similarity scores and ALINE similarity score.

word	translation	%correct	lv	aline
casa	casa	100.0	0.500	0.870
cavall	caballo	100.0	0.667	0.926
cotxe	coche	100.0	0.500	0.970

estrella	estrella	100.0	0.714	0.984
flor	flor	100.0	0.500	0.829
formiga	hormiga	100.0	0.286	0.806
girafa	jirafa	100.0	0.667	0.873
globus	globo	100.0	0.500	0.866
gorra	gorra	100.0	0.750	0.970
guitarra	guitarra	100.0	0.833	0.980
lluna	luna	100.0	0.500	0.910
patata	patata	100.0	0.667	0.967
planta	planta	100.0	0.833	0.982
tren	tren	100.0	0.750	1.000
xocolata	chocolate	100.0	0.500	0.960
boca	boca	97.0	0.750	0.970
gat	gato	97.0	0.750	0.919
moto	moto	97.0	0.750	0.970
sol	sol	97.0	0.667	1.000
tobogan	tobogan	97.0	0.714	0.973
camisa	camisa	93.9	1.000	1.000
zebra	cebra	93.9	0.400	0.830
moneda	moneda	90.9	0.500	0.967
aigua	agua	87.9	0.800	0.895
bici	bici	87.9	0.750	1.000
corona	corona	87.9	0.667	0.967
porta	puerta	87.9	0.500	0.877

tomaquet	tomate	87.9	0.429	0.806
pingui	pinguino	84.8	0.750	0.857
boto	boton	81.8	0.600	0.826
arbre	arbol	75.8	0.200	0.485
jaqueta	chaqueta	75.8	0.333	0.900
llibre	libro	75.8	0.600	0.896
tigre	tigre	75.8	0.600	0.904
plat	plato	69.7	0.800	0.941
bol	bol	57.6	0.667	1.000
ploma	pluma	57.6	0.600	0.963
bibero	biberon	54.5	0.714	0.878
pa	pan	54.5	0.667	0.741
estruc	avestruz	45.5	0.500	0.856
lloro	loro	42.4	0.500	0.910
tisores	tijeras	39.4	0.571	0.822
orella	oreja	36.4	0.200	0.661
galeta	galleta	33.3	0.333	0.927
esponja	esponja	30.3	0.286	0.832
abella	abeja	27.3	0.200	0.661
caixa	caja	27.3	0.500	0.910
pollastre	pollo	27.3	0.250	0.563
poma	manzana	27.3	0.143	0.477
fulla	hoja	24.2	0.000	0.400
nas	nariz	24.2	0.400	0.773

ovella	oveja	24.2	0.000	0.574
ull	ojo	15.2	0.000	0.226
papallona	mariposa	12.1	0.125	0.475
taronja	naranja	12.1	0.143	0.703
got	vaso	9.1	0.000	0.530
lleo	leon	9.1	0.250	0.679
samarreta	camiseta	9.1	0.250	0.690
dit	dedo	6.1	0.250	0.649
mico	mono	6.1	0.250	0.550
bolet	seta	3.0	0.000	0.502
bou	toro	3.0	0.250	0.558
entrepa	bocadillo	3.0	0.000	0.431
forquilla	tenedor	3.0	0.000	0.378
gerd	cesped	3.0	0.000	0.469
ma	mano	3.0	0.500	0.667
mirall	espejo	3.0	0.000	0.225
aixeta	grifo	0.0	0.000	0.400
amanida	ensalada	0.0	0.000	0.532
ampolla	botella	0.0	0.167	0.707
berenar	merienda	0.0	0.375	0.771
bolquer	panal	0.0	0.000	0.459
cama	pierna	0.0	0.000	0.560
cirera	cereza	0.0	0.333	0.753
colom	paloma	0.0	0.500	0.744

cuc	gusano	0.0	0.167	0.468
galleda	cubo	0.0	0.000	0.544
mussol	buho	0.0	0.200	0.550
pernil	jamon	0.0	0.000	0.374
peu	pie	0.0	0.333	0.769
raspall	cepillo	0.0	0.167	0.700
ulleres	gafas	0.0	0.167	0.407
ungla	una	0.0	0.200	0.580
vaixell	barco	0.0	0.200	0.526
onada	ola	0.0	0.200	0.422

D.2 Excluded trials

Table D.4: List of excluded trials.

word	language	participant	translation
agua	Spanish	English	water
aigua	Catalan	English	water
luna	Spanish	English	moon
lluna	Catalan	English	moon
casa	Spanish	English	house
casa	Catalan	English	house
libro	Spanish	English	book
llibre	Catalan	English	book
leche	Spanish	English	milk
toro	Spanish	English	bull
pantalon	Spanish	English	trousers
pan	Spanish	English	bread
corona	Spanish	English	crown
corona	Catalan	English	crown
sol	Spanish	English	sun
sol	Catalan	English	sun
lengua	Spanish	English	tongue/language
porc	Catalan	English	pig/pork
porc	Catalan	Spanish	pig/pork

Appendix E

jTRACE word and phoneme lists

E.1 Words

Table E.1: List of presented words (Spanish) with jTRACE transcription in brackets, length of presented words in phonemes, translation in English with jTRACE transcription in brackets, jTRACE activation for the correct answer and percentage of English participants who gave correct answer.

presented	length	correct answer	activation	participant
guitarra (gitara)	6	guitar (gItAr)	0.632	100
sandwich (sandwiS)	7	sandwich (sandwItS)	0.718	96.6
tren (trEn)	4	train (trEIn)	0.409	43.3

chaqueta (tSakEta)	7	jacket (dMak^t)	-0.01	40
ensalada (Ensalada)	8	salad (sal^d)	0.553	34.5
cebra (TEbra)	5	zebra (zibr^)	-0.01	24.1
mesa (mEsa)	4	table (tEIb^l)	-0.01	10.3
pluma (pluma)	5	feather (fDZ@)	-0.01	10.3
puerta (pwErta)	6	door (dLr)	-0.01	10.3
bici (biTi)	4	bike (baIk)	-0.01	6.9
gafas (gafas)	5	glasses (glas^z)	-0.01	6.9
pie (pjE)	3	foot (fUt)	-0.01	6.9
camiseta (kamisEta)	8	t-shirt (tis@t)	-0.01	3.3
manzana (maNTana)	7	apple (ap^l)	-0.01	3.3
calcetin (kalTEtin)	8	sock (sAk)	0.078	0
camisa (kamisa)	6	shirt (S@t)	-0.01	0
cereza (TErETa)	6	cherry (tSDri)	-0.01	0
cesta (TEsta)	5	basket (bask^t)	-0.01	0
cuchara (kutSara)	7	spoon (spun)	-0.01	0
lapiz (lapiT)	5	pencil (pDns^l)	-0.01	0
lengua (lENgwa)	6	tongue (t^N)	-0.01	0
muneca (muNEka)	6	doll (dAl)	-0.01	0
panal (paNal)	5	nappy (napi)	-0.01	0
pastel (pastEl)	6	cake (kEIk)	-0.01	0
pez (pET)	3	fish (fIS)	-0.01	0
pierna (pjErna)	6	leg (lDg)	-0.01	0

Table E.2: List of presented words (Catalan) with jTRACE transcription in brackets, length of presented words in phonemes, translation in English with jTRACE transcription in brackets, jTRACE activation for the correct answer and percentage of English participants who gave correct answer.

presented	length	correct answer	activation	participant
guitarra (gitar@)	6	guitar (gItAr)	0.623	97.1
xocolata (Sukulat@)	8	chocolate (tSLkl^t)	-0.01	97.1
girafa (Miraf@)	6	giraffe (dM@af)	-0.01	85.3
pingui (piNgwi)	6	penguin (pDNgw^n)	0.690	85.3
planta (plant@)	6	plant (plant)	0.765	82.4
esponja (@spLnM@)	7	sponge (sp^ndM)	-0.01	79.4
jaqueta (M@kDt@)	6	jacket (dMak^t)	0.496	55.9
tigre (tigr@)	5	tiger (taIg@)	-0.01	55.9
tren (trDn)	4	train (trEIn)	-0.01	35.3
flor (flL)	3	flower (flaU@)	0.049	26.5
gat (gat)	3	cat (kat)	-0.01	20.6
camisa (kamisa)	6	shirt (S@t)	-0.01	11.8
bici (bisi)	4	bike (baIk)	-0.01	8.8
plat (plat)	4	plate (plEIt)	-0.01	8.8
porta (pLrt@)	5	door (dLr)	-0.01	8.8
cotxe (kLtS@)	5	car (kAr)	-0.01	5.9
taronja (t@rLNM@)	7	orange (Lr^ndM)	-0.01	5.9

estruc (@strus)	6	ostrich (LstrItS)	-0.01	2.9
formiga (furmig@)	7	ant (ant)	-0.01	2.9
mico (miku)	4	monkey (m^Nki)	-0.01	2.9
moneda (munDZ@)	6	coin (kLIn)	-0.01	2.9
ungla (uNgl@)	5	nail (nEIl)	-0.01	2.9
aigua (ajgw@)	5	water (wLt@)	-0.01	0
aixeta (@SDt@)	5	tap (tap)	-0.004	0
amanida (@m@niZ@)	7	salad (sal^d)	-0.01	0
berenar (b@r@na)	6	snack (snak)	-0.01	0
bolet (bulDt)	5	mushroom (m^Srum)	-0.01	0
bolquer (bulkE)	5	nappy (napi)	-0.01	0
caixa (kaS@)	4	box (bAks)	-0.01	0
cirera (sirEr@)	6	cherry (tSDri)	-0.01	0
cuc (kuk)	3	worm (w@m)	-0.01	0
dit (dit)	3	finger (fINg@)	-0.01	0
entrepa (Entr@pa)	7	sandwich (sandwItS)	-0.01	0
galeta (g@lDt@)	6	biscuit (bIsk^t)	-0.01	0
gerd (MDrt)	4	grass (gras)	-0.01	0
got (gLt)	3	cup (k^p)	-0.01	0
ma (ma)	2	hand (hand)	-0.01	0
mussol (musLl)	5	owl (aUl)	-0.01	0
pernil (p@rnil)	6	ham (ham)	-0.01	0
peu (pDw)	3	foot (fUt)	-0.01	0

E.2 Phonemes

Table E.3: List of phonemes (consonants) implemented in jTRACE, matched for IPA and jTRACE transcriptions, with voicing, place of articulation and manner of articulation.

IPA	jTRACE	Voicing	Place	Manner
Strauss, Harris and Magnuson (2007)				
b	b	Voiced	Bilabial	Stop
p	p	Voiceless	Bilabial	Stop
d	d	Voiced	Alveolar	Stop
t	t	Voiceless	Alveolar	Stop
g	g	Voiced	Velar	Stop
k	k	Voiceless	Velar	Stop
l	l	Voiced	Alveolar	Lateral Appr.
ɹ	r	Voiced	Alveolar	Approximant
s	s	Voiceless	Alveolar	Fricative
ʃ	S	Voiceless	Post-alveolar	Fricative
Mayor and Plunkett (2014)				
m	m	Voiced	Bilabial	Nasal
n	n	Voiced	Alveolar	Nasal

ŋ	N	Voiced	Velar	Nasal
v	v	Voiced	Labiodental	Fricative
f	f	Voiceless	Labiodental	Fricative
ð	Z	Voiced	Dental	Fricative
θ	T	Voiceless	Dental	Fricative
ʒ	M	Voiced	Post-alveolar	Fricative
h	h	Voiceless	Glottal	Fricative
z	z	Voiced	Alveolar	Fricative
j	j	Voiced	Palatal	Approximant
w	w	Voiced	Labialised velar	Approximant

Table E.4: List of phonemes (vowel) implemented in jTRACE, matched for IPA and jTRACE transcriptions, with vowel height, backness and roundedness.

IPA	jTRACE	Height	Backness	Roundedness
Strauss, Harris and Magnuson (2007)				
ʌ	^	Low-mid	Back	Unrounded
a	a	Low	Front	Unrounded
i	i	High	Front	Unrounded
u	u	High	Back	Rounded
Mayor and Plunkett (2014)				

ɒ	Q	Low	Back	Rounded
ɑ	A	Low	Back	Unrounded
ɛ	D	Low-mid	Front	Unrounded
ɔ	L	Low-mid	Back	Rounded
ə	@	Mid	Central	
e	E	High-mid	Front	Unrounded
ɪ	I	Near-high	Front	Unrounded
ʊ	U	Near-high	Back	Rounded

Table E.5: Activated jTRACE feature units associated to each phoneme, as presented in the jTRACE graphical interface (BUR).

IPA	jTRACE	1	2	3	4	5	6	7	8	9
silence	–	0	0	0	0	0	0	0	0	1
p	p	0	0	0	0	1	0	0	0	0
b	b	0	0	0	0	1	0	0	0	0
t	t	0	0	0	0	1	0	0	0	0
d	d	0	0	0	0	1	0	0	0	0
k	k	0	0	0	0	1	0	0	0	0
g	g	0	0	0	0	1	0	0	0	0
s	s	0	0	1	0	0	0	0	0	0
ʃ	S	0	0	1	0	0	0	0	0	0
ɹ	r	0	1	0	0	0	0	0	0	0
l	l	0	1	0	0	0	0	0	0	0
w	w	0	1	0	0	0	0	0	0	0
f	f	0	0	1	0	0	0	0	0	0
θ	T	0	0	1	0	0	0	0	0	0
n	n	0	1	0	0	0	0	0	0	0
m	m	0	1	0	0	0	0	0	0	0
ð	Z	0	0	1	0	0	0	0	0	0
z	z	0	0	1	0	0	0	0	0	0
v	v	0	0	1	0	0	0	0	0	0
ʒ	M	0	0	1	0	0	0	0	0	0

j	j	0	1	0	0	0	0	0	0	0
h	h	0	0	1	0	0	0	0	0	0
ŋ	N	0	1	0	0	0	0	0	0	0
a	a	1	0	0	0	0	0	0	0	0
i	i	1	0	0	0	0	0	0	0	0
u	u	1	0	0	0	0	0	0	0	0
ʌ	^	0	1	0	0	0	0	0	0	0
ʊ	U	1	0	0	0	0	0	0	0	0
ɒ	Q	0	1	0	0	0	0	0	0	0
ə	@	0	1	0	0	0	0	0	0	0
ɪ	I	0	1	0	0	0	0	0	0	0
ɑ:	A	1	0	0	0	0	0	0	0	0
e	E	1	0	0	0	0	0	0	0	0
ɛ	D	1	0	0	0	0	0	0	0	0
ɔ	L	1	0	0	0	0	0	0	0	0

Table E.6: Activated jTRACE feature units associated to each phoneme, as presented in the jTRACE graphical interface (VOI).

IPA	jTRACE	1	2	3	4	5	6	7	8	9
silence	–	0	0	0	0	0	0	0	0	1
p	p	0	0	0	0	0	0	0	1	0
b	b	0	0	0	0	0	0	0	1	0

t	t	0	0	0	0	0	0	0	1	0
d	d	0	0	0	0	0	0	0	1	0
k	k	0	0	0	0	0	0	0	1	0
g	g	0	0	0	0	0	0	0	1	0
s	s	0	0	0	0	1	0	0	0	0
ʃ	S	0	0	0	0	1	0	0	0	0
ɹ	r	0	1	0	0	0	0	0	0	0
l	l	0	1	0	0	0	0	0	0	0
w	w	0	1	0	0	0	0	0	0	0
f	f	0	0	0	0	1	0	0	0	0
θ	T	0	0	0	0	1	0	0	0	0
n	n	0	0	1	0	0	0	0	0	0
m	m	0	0	1	0	0	0	0	0	0
ð	Z	0	0	0	0	1	0	0	0	0
z	z	0	0	0	0	1	0	0	0	0
v	v	0	0	0	0	1	0	0	0	0
ʒ	M	0	0	0	0	1	0	0	0	0
j	j	0	1	0	0	0	0	0	0	0
h	h	0	0	0	0	1	0	0	0	0
ŋ	N	0	0	1	0	0	0	0	0	0
a	a	1	0	0	0	0	0	0	0	0
i	i	1	0	0	0	0	0	0	0	0
u	u	1	0	0	0	0	0	0	0	0

ʌ	^	1	0	0	0	0	0	0	0	0
ʊ	U	1	0	0	0	0	0	0	0	0
ɒ	Q	1	0	0	0	0	0	0	0	0
ə	@	1	0	0	0	0	0	0	0	0
ɪ	I	1	0	0	0	0	0	0	0	0
ɑ:	A	1	0	0	0	0	0	0	0	0
e	E	1	0	0	0	0	0	0	0	0
ɛ	D	1	0	0	0	0	0	0	0	0
ɔ	L	1	0	0	0	0	0	0	0	0

Table E.7: Activated jTRACE feature units associated to each phoneme, as presented in the jTRACE graphical interface (GRD).

IPA	jTRACE	1	2	3	4	5	6	7	8	9
silence	–	0	0	0	0	0	0	0	0	1
p	p	0	1	0	0	0	0	0	0	0
b	b	0	1	0	0	0	0	0	0	0
t	t	0	1	0	0	0	0	0	0	0
d	d	0	1	0	0	0	0	0	0	0
k	k	0	0	0	0	0	0	1	0	0
g	g	0	0	0	0	0	0	1	0	0
s	s	0	1	0	0	0	0	0	0	0
ʃ	S	0	0	1	0	0	0	0	0	0
ɹ	r	0	0	0	0	0	0	0.5	1	0

l	l	0	0	0	0	0	0	1	0.5	0
w	w	0	1	0	0	0	0	0	0	0
f	f	0	1	0	0	0	0	0	0	0
θ	T	0	1	0	0	0	0	0	0	0
n	n	0	1	0	0	0	0	0	0	0
m	m	0	1	0	0	0	0	0	0	0
ð	Z	0	1	0	0	0	0	0	0	0
z	z	0	1	0	0	0	0	0	0	0
v	v	0	1	0	0	0	0	0	0	0
ʒ	M	0	0	1	0	0	0	0	0	0
j	j	1	0	0	0	0	0	0	0	0
h	h	0	0	0	0	1	0	0	0	0
ŋ	N	0	0	0	0	0	0	1	0	0
a	a	0	0	0	0	0	0	1	0	0
i	i	1	0	0	0	0	0	0	0	0
u	u	0	0	1	0	0	0	0	0	0
ʌ	^	0	0	0	1	0	0	0	0	0
ʊ	U	0	1	0	0	0	0	0	0	0
ɒ	Q	0	0	0	0	0	0	1	0	0
ə	@	0	0	0	1	0	0	0	0	0
ɪ	I	0	0	0	0	0	0	1	0.5	0
ɑː	A	0	0	0	0	0	0	1	0	0
e	E	0	1	0	0	0	0	0	0	0

ɛ	D	0	0	0	0	1	0	0	0	0
ɔ	L	0	0	0	0	1	0	0	0	0

Table E.8: Activated jTRACE feature units associated to each phoneme, as presented in the jTRACE graphical interface (ACU).

IPA	jTRACE	1	2	3	4	5	6	7	8	9
silence	–	0	0	0	0	0	0	0	0	1
p	p	0	0	0	0	0	0	1	0	0
b	b	0	0	0	0	0	0	1	0	0
t	t	0	1	0	0	0	0	0	0	0
d	d	0	1	0	0	0	0	0	0	0
k	k	0	0	0	0.1	0.3	1	0.3	0.1	0
g	g	0	0	0	0.1	0.3	1	0.3	0.1	0
s	s	1	0.3	0.1	0	0	0	0	0	0
ʃ	S	0	0	0.1	0.3	1	0.3	0.1	0	0
ɹ	r	0	0	0	0	0	0	1	0	0
l	l	0	0	0	0	1	0	0	0	0
w	w	0	0	0	0	0	0	1	0	0
f	f	0	0	0	0	0	1	0	0	0
θ	T	0	0	0	0	1	0	0	0	0
n	n	0	1	0	0	0	0	0	0	0
m	m	0	0	0	0	0	0	1	0	0
ð	Z	0	1	0	0	0	0	0	0	0

z	z	1	0	0	0	0	0	0	0	0
v	v	0	0	0	0	0	1	0	0	0
ʒ	M	0	0	0	0	1	0	0	0	0
j	j	1	0	0	0	0	0	0	0	0
h	h	0	0	0	0	0	0	0	1	0
ŋ	N	0	0	0	0	0	1	0	0	0
a	a	0	0	0	0	0	0.1	0.3	1	0
i	i	1	0.3	0.1	0	0	0	0	0	0
u	u	0	0	0	0	0.1	0.3	1	0.3	0
ʌ	^	0	0	0	0	0	0.1	0.3	1	0
ʊ	U	0	0	0	0	0	1	0	0	0
ɒ	Q	0	0	0	0	0	0	1	0	0
ə	@	0	0	0	0	0	0	0	1	0
ɪ	I	0	0	0	0	1	0	0	0	0
ɑː	A	0	0	0	0	0	0.3	1	0	0
e	E	0	1	0	0	0	0	0	0	0
ɛ	D	0	0	1	0	0	0	0	0	0
ɔ	L	0	0	0	0	0	0	1	0	0

Table E.9: Activated jTRACE feature units associated to each phoneme, as presented in the jTRACE graphical interface (DIF).

IPA	jTRACE	1	2	3	4	5	6	7	8	9
silence	–	0	0	0	0	0	0	0	0	1
p	p	0	0	0	0	0	0	0	1	0
b	b	0	0	0	0	0	0	0	1	0
t	t	0	0	0	0	0	0	0	1	0
d	d	0	0	0	0	0	0	0	1	0
k	k	0	0	0	0	0	0	0	1	0
g	g	0	0	0	0	0	0	0	1	0
s	s	0	0	0	1	0	0	0	0	0
ʃ	S	0	0	0	1	0	0	0	0	0
ɹ	r	0	0	0	0	1	0	0	0	0
l	l	0	0	0	0	1	0	0	0	0
w	w	0	0	1	0	0	0	0	0	0
f	f	0	0	0	0	0	1	0	0	0
θ	T	0	0	0	0	0	1	0	0	0
n	n	0	0	0	0	1	0	0	0	0
m	m	0	0	0	0	1	0	0	0	0
ð	Z	0	1	0	0	0	0	0	0	0
z	z	0	0	0	0	0	1	0	0	0
v	v	0	0	0	0	0	1	0	0	0
ʒ	M	0	0	0	0	0	1	0	0	0
j	j	0	0	1	0	0	0	0	0	0
h	h	0	0	0	0	0	1	0	0	0
ŋ	N	0	0	0	0	1	0	0	0	0

a	a	0	1	0	0	0	0	0	0	0
i	i	0	1	0	0	0	0	0	0	0
u	u	0	1	0	0	0	0	0	0	0
ʌ	^	0	1	0	0	0	0	0	0	0
ʊ	U	0	1	0	0	0	0	0	0	0
ɒ	Q	1	0	0	0	0	0	0	0	0
ə	@	0	1	0	0	0	0	0	0	0
ɪ	I	0	0	0	0	1	0	0	0	0
ɑː	A	0	1	0	0	0	0	0	0	0
e	E	0	1	0	0	0	0	0	0	0
ɛ	D	0	1	0	0	0	0	0	0	0
ɔ	L	0	1	0	0	0	0	0	0	0

Table E.10: Activated jTRACE feature units associated to each phoneme, as presented in the jTRACE graphical interface (VOC).

IPA	jTRACE	1	2	3	4	5	6	7	8	9
silence	–	0	0	0	0	0	0	0	0	1
p	p	0	0	0	0	0	0	0	1	0
b	b	0	1	0	0	0	0	0	0	0
t	t	0	0	0	0	0	0	0	1	0
d	d	0	1	0	0	0	0	0	0	0
k	k	0	0	0	0	0	0	0	1	0

g	g	0	1	0	0	0	0	0	0	0
s	s	0	0	0	0	0	0	0	1	0
ʃ	S	0	0	0	0	0	0	0	1	0
ɹ	r	1	0	0	0	0	0	0	0	0
l	l	1	0	0	0	0	0	0	0	0
w	w	1	0	0	0	0	0	0	0	0
f	f	0	0	0	0	0	0	0	1	0
θ	T	0	0	0	0	0	0	0	1	0
n	n	1	0	0	0	0	0	0	0	0
m	m	1	0	0	0	0	0	0	0	0
ð	Z	1	0	0	0	0	0	0	0	0
z	z	1	0	0	0	0	0	0	0	0
v	v	1	0	0	0	0	0	0	0	0
ʒ	M	1	0	0	0	0	0	0	0	0
j	j	1	0	0	0	0	0	0	0	0
h	h	0	0	0	0	0	0	0	1	0
ŋ	N	1	0	0	0	0	0	0	0	0

a	a	1	0	0	0	0	0	0	0	0
i	i	1	0	0	0	0	0	0	0	0
u	u	1	0	0	0	0	0	0	0	0
ʌ	^	1	0	0	0	0	0	0	0	0
ʊ	U	1	0	0	0	0	0	0	0	0
ɒ	Q	1	0	0	0	0	0	0	0	0

ə	@	1	0	0	0	0	0	0	0	0
ɪ	I	1	0	0	0	0	0	0	0	0
ɑ:	A	1	0	0	0	0	0	0	0	0
e	E	1	0	0	0	0	0	0	0	0
ɛ	D	1	0	0	0	0	0	0	0	0
ɔ	L	1	0	0	0	0	0	0	0	0

Table E.11: Activated jTRACE feature units associated to each phoneme, as presented in the jTRACE graphical interface (POW).

IPA	jTRACE	1	2	3	4	5	6	7	8	9
silence	–	0	0	0	0	0	0	0	0	1
p	p	1	0.2	0	0	0	0	0	0	0
b	b	0.2	1	0	0	0	0	0	0	0
t	t	0	0	1	0.2	0	0	0	0	0
d	d	0	0	0.2	1	0	0	0	0	0
k	k	0	0	0	0	1	0.2	0	0	0
g	g	0	0	0	0	0.2	1	0	0	0
s	s	0	0	0	0	0	0	0	0	0
ʃ	S	0	0	0	0	0	0	0	0	0
ɹ	r	0	0	0	0	0	0	0	0	0
l	l	0	0	0	0	0	0	0	0	0
w	w	0	0	0	0	0	0	0	0	0
f	f	0	0	0	0	0	0	0	0	0

θ	T	0	0	0	0	0	0	0	0	0
n	n	0	0	0	0	0	0	0	0	0
m	m	0	0	0	0	0	0	0	0	0
ð	Z	0	0	0	0	0	0	0	0	0
z	z	0	0	0	0	0	0	0	0	0
v	v	0	0	0	0	0	0	0	0	0
ʒ	M	0	0	0	0	0	0	0	0	0
j	j	0	0	0	0	0	0	0	0	0
h	h	0	0	0	0	0	0	0	0	0
ŋ	N	0	0	0	0	0	0	0	0	0
a	a	0	0	0	0	0	0	0	0	0
i	i	0	0	0	0	0	0	0	0	0
u	u	0	0	0	0	0	0	0	0	0
ʌ	^	0	0	0	0	0	0	0	0	0
ʊ	U	0	0	0	0	0	0	0	0	0
ɒ	Q	0	0	0	0	0	0	0	0	0
ə	@	0	0	0	0	0	0	0	0	0
ɪ	I	0	0	0	0	0	0	0	0	0
ɑː	A	0	0	0	0	0	0	0	0	0
e	E	0	0	0	0	0	0	0	0	0
ɛ	D	0	0	0	0	0	0	0	0	0
ɔ	L	0	0	0	0	0	0	0	0	0

References

Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child development*, *72*(2), 416–430.

Allen, D. (2019). Cognate frequency and assessment of second language lexical knowledge. *International Journal of Bilingualism*, *23*(5), 1121–1136.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *science*, *286*(5439), 509–512.

Barry, H., & Paxson, L. M. (1971). Infancy and early childhood: Cross-cultural codes 2. *Ethnology*, *10*(4), 466–508.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PloS one*, *6*(5), e19348.

Benavides-Varela, S., Hochmann, J.-R., Macagno, F., Nespor, M., & Mehler, J. (2012). Newborn's brain activity signals the origin of word memories. *Proceedings of the National Academy of Sciences*, *109*(44), 17908–17913.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by english-speaking adults and infants. *Journal of experimental psychology: human perception and performance*, *14*(3), 345.

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

Bialystok, E., Luk, G., Peets, K. F., & Sujin, Y. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and cognition*, *13*(4), 525–531.

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental science*, *7*(3), 325–339.

Bilson, S., Yoshida, H., Tran, C. D., Woods, E. A., & Hills, T. T. (2015). Semantic facilitation in bilingual first language acquisition. *Cognition*, *140*, 122–134.

Blumenfeld, H. K., Bobb, S. C., & Marian, V. (2016). The role of language proficiency, cognate status and word frequency in the assessment of spanish–english bilinguals' verbal fluency. *International journal of speech-language pathology*, *18*(2), 190–201.

Blythe, R. A., Smith, K., & Smith, A. D. (2010). Learning times for large lexicons through cross-situational learning. *Cognitive Science*, *34*(4), 620–642.

Bond, Z. S., & Small, L. H. (1983). Voicing, vowel, and stress mispronunciations in continuous speech. *Perception & Psychophysics*, *34*(5), 470–474.

Bosch, L., & Ramon-Casas, M. (2014). First translation equivalents in bilingual toddlers' expressive vocabulary: Does form similarity matter? *International Journal of Behavioral Development*, *38*(4), 317–322.

Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, *2*(1), 29–49.

Bosma, E., Blom, E., Hoekstra, E., & Versloot, A. (2019). A longitudinal study on the gradual cognate facilitation effect in bilingual children's frisian receptive vocabulary. *International Journal of Bilingual Education and Bilingualism*, *22*(4), 371–385.

Braginsky, M., Sanchez, A., & Yurovsky, D. (2020). childesr: Accessing the 'childes' database [Computer software manual]. Retrieved from `https://github.com/langcog/childesr` (R package version 0.1.2)

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind*, *3*, 52–67.

Brenders, P., Van Hell, J. G., & Dijkstra, T. (2011). Word recognition in child second language learners: Evidence from cognates and false friends. *Journal of experimental child psychology*, *109*(4), 383–396.

Brownell, R. (2000). *Expressive one-word picture vocabulary test: Manual.* Academic Therapy Publications.

Buac, M., Gross, M., & Kaushanskaya, M. (2014). The role of primary caregiver vocabulary knowledge in the development of bilingual children's vocabulary skills. *Journal of Speech, Language, and Hearing Research*, *57*(5), 1804–1816.

Bultena, S., Dijkstra, T., & Van Hell, J. G. (2014). Cognate effects in sentence context depend on word class, l2 proficiency, and task. *Quarterly Journal of Experimental Psychology*, *67*(6), 1214–1241.

Byers-Heinlein, K. (2013). Parental language mixing: Its measurement and the relation of mixed input to young bilingual children's vocabulary size. *Bilingualism: Language and Cognition*, *16*(1), 32–48.

Byers-Heinlein, K., & Werker, J. F. (2013). Lexicon structure and the disambiguation of novel words: Evidence from bilingual infants. *Cognition*, *128*(3), 407–416.

Cattani, A., Abbot-Smith, K., Farag, R., Krott, A., Arreckx, F., Dennis, I., & Floccia, C. (2014). How much exposure to english is necessary for a bilingual toddler to perform like a monolingual peer in language tests? *International journal of language & communication disorders*, *49*(6), 649–671.

Chang, L. M., & Deák, G. O. (2020). Adjacent and non-adjacent word contexts both predict age of acquisition of english words: A distributional corpus analysis of child-directed speech. *Cognitive Science*, *44*(11), e12899.

Chow, J., Aimola Davies, A. M., Fuentes, L. J., & Plunkett, K. (2016). Backward semantic inhibition in toddlers. *Psychological Science*, *27*(10), 1312–1320.

Christoffels, I. K., De Groot, A. M., & Kroll, J. F. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language*, *54*(3), 324–345.

Cole, R. A., & Jakimik, J. (1980). How are syllables used to recognize words? *The Journal of the Acoustical Society of America*, *67*(3), 965–970.

Cole, R. A., Jakimik, J., & Cooper, W. E. (1978). Perceptibility of phonetic features in fluent speech. *The Journal of the Acoustical Society of America*, *64*(1), 44–56.

Connine, C. M. (1994). Vertical and horizontal similarity in spoken word recognition. *Perspectives on sentence processing*, 107–120.

Connine, C. M., Blasko, D. G., & Wang, J. (1994). Vertical similarity in spoken word recognition: Multiple lexical activation, individual differences, and the role of sentence context. *Perception & Psychophysics*, *56*(6), 624–636.

Cook, S. V., Pandža, N. B., Lancaster, A. K., & Gor, K. (2016). Fuzzy nonnative phonolexical representations lead to fuzzy form-to-meaning mappings.

Frontiers in Psychology, *7*, 1345.

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in english: Evidence from native and non-native listeners. *Language and speech*, *45*(3), 207–228.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1283.

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and language*, *41*(3), 365–397.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2006). Acquiring an artificial lexicon: Segment type and order information in early lexical entries. *Journal of Memory and Language*, *54*(1), 1–19.

Cristia, A. (2022). A systematic review suggests marked differences in the prevalence of infant-directed vocalization across groups of populations. *Developmental Science*, e13265.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, *Complex Systems*, 1695. Retrieved from `http://igraph.org`

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The "small world of words" english word association norms for over 12,000 cue words. *Behavior research methods*, *51*(3), 987–1006.

De Houwer, A., Bornstein, M. H., & De Coster, S. (2006). Early understanding of two words for the same thing: A cdi study of lexical comprehension in infant bilinguals. *International Journal of Bilingualism*, *10*(3), 331–347.

De Houwer, A., Bornstein, M. H., & Putnick, D. L. (2014). A bilingual–monolingual comparison of young children's vocabulary size: Evidence from comprehension and production. *Applied psycholinguistics*, *35*(6), 1189–1211.

Desroches, A. S., Friesen, D. C., Teles, M., Korade, C. A., & Forest, E. W. (2022). The dynamics of spoken word recognition in bilinguals. *Bilingualism: Language and Cognition*, 1–6.

Dobel, C., Lagemann, L., & Zwitserlood, P. (2009). Non-native phonemes in adult word learning: evidence from the n400m. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3697–3709.

Downey, S. S., Hallmark, B., Cox, M. P., Norquest, P., & Lansing, J. S. (2008). Computational feature-sensitive reconstruction of language relationships: Developing the aline distance for comparative historical linguistic reconstruction. *Journal of Quantitative Linguistics*, *15*(4), 340–369.

Downey, S. S., Sun, G., & Norquest, P. (2017). aliner: an r package for optimizing feature-weighted alignments and linguistic distances. *The R Journal*, *9*(1), 138–152. Retrieved from `https://journal.r-project.org/archive/2017/RJ-2017-005/index.html` (R Package Version 1.1.4)

Dufour, S., Nguyen, N., & Frauenfelder, U. H. (2007). The perception of phonemic contrasts in a non-native dialect. *The Journal of the Acoustical Society of America*, *121*(4), EL131–EL136.

Dufour, S., & Peereman, R. (2003a). Inhibitory priming effects in auditory word recognition: When the target's competitors conflict with the prime word. *Cognition*, *88*(3), B33–B44.

Dufour, S., & Peereman, R. (2003b). Lexical competition in phonological priming:

Assessing the role of phonological match and mismatch lengths between primes and targets. *Memory & Cognition*, *31*(8), 1271–1283.

Duñabeitia, J. A., Ivaz, L., & Casaponsa, A. (2016). Developmental changes associated with cross-language similarity in bilingual children. *Journal of Cognitive Psychology*, *28*(1), 16–31.

Dunn, L., & Dunn, L. (1997). *Ppvt: Peabody picture vocabulary test—third edition (ppvt-iii).* Pearson Assessments, Bloomington, MN.

Dunn, L. M., & Dunn, D. M. (2009). *The british picture vocabulary scale.* GL Assessment Limited.

Erdős, P., & Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, *5*(1), 17–60.

Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and speech*, *46*(2-3), 245–264.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., ... Stiles, J. (1994). Variability in early communicative development. *Monographs of the society for research in child development*, i–185.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL quarterly*, *39*(3), 399–423.

Floccia, C., Sambrook, T. D., Delle Luche, C., Kwok, R., Goslin, J., White, L., ... Plunkett, K. (2018). Vocabulary of 2-year-olds learning english and an additional language: norms and effects of linguistic distance. *Monographs of the Society for Research in Child Development*, *83*(1), 1–135.

Foursha-Stevenson, C., Schembri, T., Nicoladis, E., & Eriksen, C. (2017). The influence of child-directed speech on word learning and comprehension. *Journal of psycholinguistic research*, *46*(2), 329–343.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of child language*, *44*(3), 677.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological science*, *20*(5), 578–585.

Gámez, P. B., Palermo, F., Perry, J. S., & Galindo, M. (2022). Spanish-english bilingual toddlers' vocabulary skills: The role of caregiver language input and warmth. *Developmental Science*, e13308.

Gampe, A., Quick, A. E., & Daum, M. M. (2021). Does linguistic similarity affect early simultaneous bilingual language acquisition? *Journal of Language Contact*, *13*(3), 482–500.

Garcia-Castro, G. (2021). jtracer [Computer software manual]. Retrieved from `https://github.com/gongcastro/jtracer`

Gershkoff-Stowe, L., & Hahn, E. R. (2013). Word comprehension and production asymmetries in children and adults. *Journal of experimental child psychology*, *114*(4), 489–509.

Gillen, N. A., Siow, S., Lepadatu, I., Sucevic, J., Plunkett, K., & Duta, M. (2021). Tapping into the potential of remote developmental research: introducing the oxfordbabylab app.

Gleitman, L. (1990). The structural sources of verb meanings. *Language acquisition*, *1*(1), 3–55.

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? parental input and the acquisition of vocabulary. *Journal of child language*, *35*(3), 515.

Green, J. R., Moore, C. A., Higashikawa, M., & Steeve, R. W. (2000). The physiologic development of speech motor control: Lip and jaw coordination. *Journal of Speech, Language, and Hearing Research*, *43*(1), 239–255.

Green, J. R., Nip, I. S., Maassen, B., & Van Lieshout, P. (2010). Some organization principles in early speech development. *Speech Motor Control: New developments in basic and applied research*, *10*, 171–188.

Gross, M., Buac, M., & Kaushanskaya, M. (2014). Conceptual scoring of receptive and expressive vocabulary measures in simultaneous and sequential bilingual children. *American Journal of Speech-Language Pathology*, *23*(4), 574–586.

Hall, C. J. (2002). The automatic cognate form assumption: Evidence for the parasitic model of vocabulary development. *International Review of Applied Linguistics in Language Teaching*, *40*(2). doi: 10.1515/iral.2002.008

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a british communicative development inventory. *Journal of child language*, *27*(3), 689–705.

Hammer, C. S., Komaroff, E., Rodriguez, B. L., Lopez, L. M., Scarpino, S. E., & Goldstein, B. (2012). Predicting spanish–english bilingual children's language abilities.

Hart, B., & Risley, T. R. (1992). American parenting of language-learning children: Persisting differences in family-child interactions observed in natural home environments. *Developmental psychology*, *28*(6), 1096.

Hashemikamangar, S. S., Bakouie, F., & Gharibzadeh, S. (2020). Children semantic network growth: A graph theory analysis. In *2020 27th national and 5th international iranian conference on biomedical engineering (icbme)* (pp. 318–321).

Hidaka, S. (2013). A computational model associating learning process, word attributes, and age of acquisition. *PLOS one*, *8*(11), e76242.

Hills, T. (2013). The company that words keep: comparing the statistical structure of child-versus adult-directed language. *Journal of child language*, *40*(3), 586–604.

Hills, T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of memory and language*, *63*(3), 259–273.

Hills, T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? *Psychological science*, *20*(6), 729–739.

Hipfner-Boucher, K., Pasquarella, A., Chen, X., & Deacon, S. H. (2016). Cognate awareness in french immersion students: Contributions to grade 2 reading comprehension. *Scientific Studies of Reading*, *20*(5), 389–400.

Hochmann, J.-R., Benavides-Varela, S., Fló, A., Nespor, M., & Mehler, J. (2018). Bias for vocalic over consonantal information in 6-month-olds. *Infancy*, *23*(1), 136–151.

Hochmann, J.-R., Benavides-Varela, S., Nespor, M., & Mehler, J. (2011). Consonants and vowels: different roles in early language acquisition. *Developmental science*, *14*(6), 1445–1458.

Hochmann, J.-R., Endress, A. D., & Mehler, J. (2010). Word frequency as a cue for identifying function words in infancy. *Cognition*, *115*(3), 444–457.

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of child language*, *39*(1), 1.

Hoff, E., Core, C., & Shanks, K. F. (2020). The quality of child-directed speech depends on the speaker's language proficiency. *Journal of Child Language*, *47*(1), 132–145.

Hoff, E., & Ribot, K. M. (2017). Language growth in english monolingual and spanish-english bilingual children from 2.5 to 5 years. *The Journal of pediatrics*, *190*, 241–245.

Højen, A., & Nazzi, T. (2016). Vowel bias in danish word-learning: Processing biases are language-specific. *Developmental Science*, *19*(1), 41–49.

Hoshino, N., & Kroll, J. F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition*, *106*(1), 501–511.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Howard, E. R., Páez, M. M., August, D. L., Barr, C. D., Kenyon, D., & Malabonga, V. (2014). The importance of ses, home and school language and literacy practices, and oral vocabulary in bilingual children's english reading development. *Bilingual Research Journal*, *37*(2), 120–141.

Hutchison, K. A., Balota, D. A., Cortese, M. J., & Watson, J. M. (2008). Predicting semantic priming at the item level. *Quarterly Journal of experimental psychology*, *61*(7), 1036–1066.

Iuzzini-Seigel, J., Hogan, T. P., Rong, P., & Green, J. R. (2015). Longitudinal development of speech motor control: Motor and linguistic factors. *Journal of motor learning and development*, *3*(1), 53.

Kachergis, G., Shiffrin, R., & Yu, C. (2009). Frequency and contextual diversity effects in cross-situational word learning. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 31).

Karaminis, T., & Scharenborg, O. (2018). The effects of background noise on native and non-native spoken-word recognition: a computational modelling approach. In *The 40th annual conference of the cognitive science society* (pp. 1902–

1907).

Kartushina, N., Mani, N., Aktan-Erciyes, A., Alaslani, K., Aldrich, N. J., Almohammadi, A., ... others (2021). Covid-19 first lockdown as a unique window into language acquisition: What you do (with your child) matters.

Kenett, Y. N., Levi, E., Anaki, D., & Faust, M. (2017). The semantic distance task: Quantifying semantic distance with semantic network path length. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(9), 1470.

Kondrak, G. (2003). Phonetic alignment and similarity. *Computers and the Humanities*, *37*(3), 273–291.

Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, *107*(1), 54–81.

Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological science*, *19*(4), 332–338.

Kremin, L. V., & Byers-Heinlein, K. (2020). Why not both? rethinking categorical and continuous approaches to bilingualism.

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second language research*, *18*(2), 137–171.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of memory and language*, *33*(2), 149–174.

Kunert, R., Fernández, R., & Zuidema, W. (2011). Adaptation in child directed speech: Evidence from corpora. *Proc. SemDial*, 112–119.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–

26. doi: 10.18637/jss.v082.i13

Landau, B., & Gleitman, L. R. (2009). *Language and experience: Evidence from the blind child* (Vol. 8). Harvard University Press.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive development*, *3*(3), 299–321.

Legacy, J., Reider, J., Crivello, C., Kuzyk, O., Friend, M., Zesiger, P., & Poulin-Dubois, D. (2017). Dog or chien? translation equivalents in the receptive and expressive vocabularies of young french–english bilinguals. *Journal of child language*, *44*(4), 881–904.

Levelt, W. J., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological review*, *98*(1), 122.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady* (Vol. 10, pp. 707–710).

Li, H., & Fang, A. C. (2011). Word frequency of the childes corpus: Another perspective of child language features. *ICAME Journal*, *35*, 95–116.

Li, P., & MacWhinney, B. (2002). Patpho: A phonological pattern generator for neural networks. *Behavior Research Methods, Instruments, & Computers*, *34*(3), 408–415.

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic bulletin & review*, *7*(4), 618–630.

Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics*, *39*(3), 155–158.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and hearing*, *19*(1), 1.

MacWhinney, B. (2000). *The childes project: Tools for analyzing talk. transcription format and programs* (Vol. 1). Psychology Press.

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, *5*(9), 1315–1316.

Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language*, *57*(2), 252–272.

Mani, N., & Plunkett, K. (2010a). In the infant's mind's ear: Evidence for implicit naming in 18-month-olds. *Psychological science*, *21*(7), 908–913.

Mani, N., & Plunkett, K. (2010b). Twelve-month-olds know their cups from their keps and tups. *Infancy*, *15*(5), 445–470.

Mani, N., & Plunkett, K. (2011). Phonological priming and cohort effects in toddlers. *Cognition*, *121*(2), 196–206.

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental science*, *11*(3), F9–F16.

Marchman, V. A., & Martínez-Sussmann, C. (2002). Concurrent validity of caregiver/parent report measures of language for children who are learning both english and spanish. *Journal of Speech, Language, and Hearing Research*, *45*, 983–997.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive psychology*, *20*(2), 121–157.

Marks, E. A., Moates, D. R., Bond, Z., & Vasquez, L. (2002). Vowel mutability: The case of monolingual spanish listeners and bilingual spanish-english listeners. *Southwest Journal of Linguistics*, *21*(2), 73–100.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1-2), 71–102.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive psychology*, *10*(1), 29–63.

Mayor, J., & Plunkett, K. (2014). Infant word recognition: Insights from trace simulations. *Journal of Memory and Language*, *71*(1), 89–123.

McClelland, J. L., & Elman, J. L. (1986). The trace model of speech perception. *Cognitive psychology*, *18*(1), 1–86.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological review*, *119*(4), 831.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior research methods*, *37*(4), 547–559.

Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. W. (2007). Early activation of object names in visual search. *Psychonomic Bulletin & Review*, *14*(4), 710–716.

Meyer, A. S., & Damian, M. F. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory & cognition*, *35*(3), 494–503.

Miller, G. A. (1998). *Wordnet: An electronic lexical database*. MIT press.

Moon, K.-W. (2021). ggiraphextra: Make interactive 'ggplot2'. extension to 'ggplot2' and 'ggiraph' [Computer software manual]. Retrieved from `https://github.com/cardiomoon/ggiraphExtra` (R package version 0.3.0)

Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature neuroscience*, *6*(2), 182–189.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of child language*, *17*(2), 357–374.

Naigles, L. R., & Hoff-Ginsberg, E. (1998). Why are some verbs learned before other verbs? effects of input frequency and structure on children's early verb use. *Journal of child language*, *25*(1), 95–120.

Nelson, D. L. (1999). The university of south florida word association norms. *http://w3. usf. edu/FreeAssociation*.

Nelson, D. L., McEvoy, C. L., & Dennis, S. (2000). What is free association and what does it measure? *Memory & cognition*, *28*(6), 887–899.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The university of south florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407.

Nishibayashi, L.-L., & Nazzi, T. (2016). Vowels, then consonants: Early bias switch in recognizing segmented word forms. *Cognition*, *155*, 188–203.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189–234.

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1209.

Office for National Statistics. (2022). *Language, england and wales: Census 2021.* Retrieved from `https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/language/bulletins/languageenglandandwales/census2021` (Released 29 November 2022)

Orena, A. J., Byers-Heinlein, K., & Polka, L. (2020). What do bilingual infants actually hear? evaluating measures of language input to bilingual-learning 10-month-olds. *Developmental science*, *23*(2), e12901.

Oshima-Takane, Y., Goodz, E., & Derevensky, J. L. (1996). Birth order effects on early language development: Do secondborn children learn from overheard speech? *Child development*, *67*(2), 621–634.

Otwinowska, A., & Szewczyk, J. M. (2017). The more similar the better? factors in learning cognates, false cognates and non-cognate words. *International Journal of Bilingual Education and Bilingualism*.

O'Toole, C., Gatt, D., Hickey, T. M., Miękisz, A., Haman, E., Armon-Lotem, S., ... Kern, S. (2017). Parent report of early lexical production in bilingual children: a cross-linguistic cdi comparison. *International Journal of Bilingual Education and Bilingualism*, *20*(2), 124–145.

O'Toole, C., & Hickey, T. M. (2017). Bilingual language acquisition in a minority context: using the irish–english communicative development inventory to track acquisition of an endangered language. *International Journal of Bilingual Education and Bilingualism*, *20*(2), 146–162.

Pearson, B. Z., & Fernández, S. C. (1994). Patterns of interaction in the lexical growth in two languages of bilingual infants and toddlers. *Language learning*, *44*(4), 617–653.

Pearson, B. Z., Fernández, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied psycholinguistics*, *18*(1), 41–58.

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language* *learning*, *43*(1), 93–120.

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1995). Cross-language synonyms in the lexicons of bilingual infants: One language or two? *Journal of child language*, *22*, 345.

Peirce, J. W. (2007). Psychopy—psychophysics software in python. *Journal of neuroscience methods*, *162*(1-2), 8–13.

Pérez, A. M., Peña, E. D., & Bedore, L. M. (2010). Cognates facilitate word recognition in young spanish-english bilinguals' test performance. *Early Childhood Services (San Diego, Calif.)*, *4*(1), 55.

Pinker, S. (1994). How could a child use verb syntax to learn verb semantics? *Lingua*, *92*, 377–410.

Place, S., & Hoff, E. (2011). Properties of dual language exposure that influence 2-year-olds' bilingual proficiency. *Child development*, *82*(6), 1834–1849.

Poarch, G. J., & Van Hell, J. G. (2012). Cross-language activation in children's speech production: Evidence from second language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, *111*(3), 419–438.

Poulin-Dubois, D., Bialystok, E., Blaye, A., Polonia, A., & Yott, J. (2013). Lexical access and vocabulary development in very young bilinguals. *International Journal of Bilingualism*, *17*(1), 57–70.

Prolific. (2014). `https://www.prolific.co`. (Accessed: 2020-05-07)

Qualtrics. (2022). Qualtrics [Computer software manual]. Provo, Utah, USA. Retrieved from `https://www.qualtrics.com`

Quine, W. V. O. (1960). *Translation and meaning*. Cambridge, MA: MIT Press.

R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Rinker, T. W. (2018). textstem: Tools for stemming and lemmatizing text [Computer software manual]. Buffalo, New York. Retrieved from `http://github.com/trinker/textstem` (R package version 0.1.4)

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). proc: an open-source package for r and s+ to analyze and compare roc curves. *BMC Bioinformatics*, *12*, 77.

Robinson Anthony, J. J., Blumenfeld, H. K., Potapova, I., & Pruitt-Lord, S. L. (2020). Language dominance predicts cognate effects and metalinguistic awareness in preschool bilinguals. *International Journal of Bilingual Education and Bilingualism*, 1–20.

Roget, P. M. (1911). *Roget's thesaurus of english words and phrases...* TY Crowell Company.

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of child language*, *35*(1), 185–205.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child development*, *83*(5), 1762–1774.

Saiegh-Haddad, E., & Haj, L. (2018). Does phonological distance impact quality of phonological representations? evidence from arabic diglossia. *Journal of child language*, *45*(6), 1377–1399.

Schepens, J., Dijkstra, T., Grootjen, F., & Van Heuven, W. J. (2013). Cross-language distributions of high frequency and phonetically similar cognates. *PloS one*, *8*(5), e63006.

Shaoul, C., & Westbury, C. (2006). Word frequency effects in high-dimensional co-occurrence models: A new approach. *Behavior Research Methods*, *38*(2), 190–195.

Sherkina-Lieber, M. (2004). The cognate facilitation effect in bilingual speech processing; the case of russian-english bilingualism. *Cahiers linguistics d'Ottawa*, *32*, 108–121.

Shneidman, L. A., & Goldin-Meadow, S. (2012). Language input and acquisition in a mayan village: How important is directed speech? *Developmental science*, *15*(5), 659–673.

Shook, A., & Marian, V. (2013). The bilingual language interaction network for comprehension of speech. *Bilingualism (Cambridge, England)*, *16*(2).

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*(1-2), 39–91.

Slowiaczek, L. M., & Hamburger, M. (1992). Prelexical facilitation and lexical interference in auditory word recognition. *Journal of experimental psychology: Learning, memory, and cognition*, *18*(6), 1239.

Smith, K., Smith, A. D., & Blythe, R. A. (2011). Cross-situational learning: An experimental study of word-learning mechanisms. *Cognitive Science*, *35*(3), 480–498.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.

Smithson, L., Paradis, J., & Nicoladis, E. (2014). Bilingualism and receptive vocabulary achievement: Could sociocultural context make a difference? *Bilingualism: language and Cognition*, *17*(4), 810–821.

Sosa, A. V., & Stoel-Gammon, C. (2012). Lexical and phonological effects in early word production. *J Speech Lang Hear Res.*, *55*(2), 596–608. doi: 10.1044/1092-

4388

Sperber, J. F., Hart, E. R., Troller-Renfree, S. V., Watts, T. W., & Noble, K. G. (2022). The effect of the covid-19 pandemic on infant development and maternal mental health in the first 2 years of life. *Infancy*.

Spinelli, E., Segui, J., & Radeau, M. (2001). Phonological priming in spoken word recognition with bisyllabic targets. *Language and cognitive processes*, *16*(4), 367–392.

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological science*, *10*(3), 281–284.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive science*, *29*(1), 41–78.

Storkel, H. L. (2004). Do children acquire dense neighborhoods? an investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics*, *25*(2), 201–221.

Strand, S., Malmberg, L., & Hall, J. (2015). English as an additional language (eal) and educational achievement in england: An analysis of the national pupil database.

Strauss, T. J., Harris, H. D., & Magnuson, J. S. (2007). jtrace: A reimplementation and extension of the trace model of speech perception and spoken word recognition. *Behavior Research Methods*, *39*(1), 19–30.

Styles, S., & Plunkett, K. (2009). What is ‘word understanding’for the parent of a one-year-old? matching the difficulty of a lexical comprehension task to parental cdi report. *Journal of child language*, *36*(4), 895–908.

Suanda, S. H., Mugwanya, N., & Namy, L. L. (2014). Cross-situational statistical word learning in young children. *Journal of experimental child psychology*, *126*, 395–411.

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological science*, *13*(5), 480–484.

Swinney, D. A. (1979). Lexical access during sentence comprehension:(re) consideration of context effects. *Journal of verbal learning and verbal behavior*, *18*(6), 645–659.

Szagun, G., Stumper, B., & Schramm, A. (2009). *Fragebogen zur frühkindlichen sprachentwicklung (frakis) und frakis-k (kurzform).* Frankfurt: Pearson Assessment.

Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First language*, *4*(12), 197–211.

Tonzar, C., Lotto, L., & Job, R. (2009). L2 vocabulary acquisition in children: Effects of learning method and cognate status. *Language Learning*, *59*(3), 623–646.

Vagh, S. B., Pan, B. A., & Mancilla-Martinez, J. (2009). Measuring growth in bilingual and monolingual children's english productive vocabulary development: The utility of combining parent and teacher report. *Child Development*, *80*(5), 1545–1563.

van der Loo, M. (2014). The stringdist package for approximate string matching. *The R Journal*, *6*, 111-122. Retrieved from `https://CRAN.R-project.org/package=stringdist`

Van Assche, E., Brysbaert, M., & Duyck, W. (2020). Bilingual lexical access. *Bilingual lexical ambiguity resolution*, 42–67.

Van den Noort, M., Struys, E., Bosch, P., Jaswetz, L., Perriard, B., Yeo, S., ... Lim, S.

(2019). Does the bilingual advantage in cognitive control exist and if so, what are its modulating factors? a systematic review. *Behavioral Sciences*, *9*(3), 27.

Van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *The Quarterly Journal of Experimental Psychology Section A*, *58*(2), 251–273.

Van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic bulletin & review*, *9*(4), 780–789.

Van Hell, J. G., & Tanner, D. (2012). Second language proficiency and cross-language lexical activation. *Language Learning*, *62*, 148–171.

van Heugten, M., Paquette-Smith, M., Krieger, D. R., & Johnson, E. K. (2018). Infants' recognition of foreign-accented words: Flexible yet precise signal-to-word mapping strategies. *Journal of Memory and Language*, *100*, 51–60.

Van Ooijen, B. (1996). Vowel mutability and lexical selection in english: Evidence from a word reconstruction task. *Memory & Cognition*, *24*(5), 573–583.

Vitevitch, M. S., Stamer, M. K., & Sereno, J. A. (2008). Word length and lexical competition: Longer is the same as shorter. *Language and Speech*, *51*(4), 361–383.

Von Holzen, K., Fennell, C. T., & Mani, N. (2019). The impact of cross-language phonological overlap on bilingual and monolingual toddlers' word recognition. *Bilingualism: Language and Cognition*, *22*(3), 476–499.

Von Holzen, K., & Mani, N. (2012). Language nonselective lexical access in bilingual toddlers. *Journal of Experimental Child Psychology*, *113*(4), 569–586.

Von Holzen, K., Nishibayashi, L.-L., & Nazzi, T. (2018). Consonant and vowel processing in word form segmentation: An infant erp study. *Brain sciences*, *8*(2), 24.

Warren-Leubecker, A., & Bohannon III, J. N. (1984). Intonation patterns in child-directed speech: Mother-father differences. *Child Development*, 1379–1385.

Weisner, T. S., Gallimore, R., Bacon, M. K., Barry III, H., Bell, C., Novaes, S. C., ... others (1977). My brother's keeper: Child and sibling caretaking [and comments and reply]. *Current anthropology*, *18*(2), 169–190.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development*, *7*(1), 49–63.

Wettler, M., Rapp, R., & Sedlmeier, P. (2005). Free word associations correspond to contiguities between words in texts. *Journal of Quantitative Linguistics*, *12*(2-3), 111–122.

White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, *59*(1), 114–132.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as bayesian inference. *Psychological review*, *114*(2), 245.

Yan, S., & Nicoladis, E. (2009). Finding le mot juste: Differences between bilingual and monolingual children's lexical access in comprehension and production. *Bilingualism: Language and Cognition*, *12*(3), 323–335.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological science*, *18*(5), 414–420.

Yurovsky, D., Smith, L. B., & Yu, C. (2013). Statistical word learning at scale: The baby's view is better. *Developmental science*, *16*(6), 959–966.

Zink, I., & Lejaegere, M. (2002). N-cdi woordenlijst [n-cdi word list]. *Leusden: Acco.*

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to* *human ecology*. Addison-Wesley Press.