UNIVERSITY of HULL

Mechanisms underpinning the gestural facilitation of second language word learning: an investigation through speeded and un-speeded tasks

being a thesis submitted in fulfilment of the

requirements for the degree of

Doctor of Philosophy

in

Psychology

at the University of Hull

by

Erin Minton-Branfoot, BSc, MSc

April 2024

Table of Contents

Acknowledgements

Firstly, I would like to thank my supervisors, Dr Henning Holle and Dr Richard O'Connor for all their support and guidance during my PhD journey. I have learnt a great deal under their wing and I am extremely grateful for all the time and effort they have put in.

A special thanks to the White Rose consortium for their funding of my research and PhD.

I would also like to thank all of my family and friends for their unwavering support.

Thank you to the volunteer actors and actresses who spent time being recorded, in order to provide the aural and visual material needed for my research. Also, thank you to the participants who took part in my studies, without whom my research would not be possible.

Lastly, thank you to all of my peers in the PhD Psychology office. When I started my PhD, I had little interaction with other PhD students due to Covid. But gradually, as we returned to campus and more students joined the office, I have gained a cohort of supportive friends.

Mechanisms underpinning the gestural facilitation of second language word learning: an investigation through speeded and un-speeded tasks

Erin Minton-Branfoot

Abstract

Vocabulary learning is one of the many challenges faced by second language learners, particularly when full immersion is not available. Iconic gesture cues have been found to provide benefits in such learning. However, the full extent of the gesture advantage has not been investigated, with gestures rarely tested as the sole cue in learning, and limited behavioural outcome measures being implemented that are unable to measure if the gesture benefits can extend to more automatic processing conditions. The series of experiments presented in this thesis aimed to test this using a new implicit, speeded task and an explicit, speeded task. Additionally, a number of proposed mechanisms that may underline the gesture advantage were investigated by systematically manipulating the cues provided during learning.

Studies 1-5 demonstrated that the gestural advantage persists when presented as the only cue during learning, but that the benefits are stronger when presented alongside an L1 translation. This may be a result of an extra level of disambiguation and the presence of an additional cue which provides encoding into the verbal store. However, the implicit, speeded task did not produce consistent results, raising questions as to whether the inclusion of these two components of automatic processing created a task that was too demanding to allow any early direct semantic effects to be detectable. Study 6 therefore implemented an explicit, speeded task to determine if the gestural advantage could extend beyond controlled retrieval. Overall, this series of studies showed that when the number of cues and level of disambiguating information provided was controlled for across conditions, the gesture-based learning method still displayed greater learning in both the explicit un-speeded task (Study 4) and the speeded task (Study 6). This research further demonstrates the robustness of the gestural advantage, with the semantic learning benefits evident in early L2 learners under rapid processing conditions. A number of factors were identified as contributing to this advantage, including the central concept that gestures have specialist, privileged access to some form of motor traces or action representations that are not provided by other cues. This research further highlights the benefits and importance of incorporating gesture-based learning methods into second language vocabulary learning, in order to develop strong conceptual links early on that are evident under automatic-like processing conditions.

INTRODUCTION

Within the following introduction, I cover several topics relating to the research area of gesture cues in second language learning encompassed in this thesis. I first discuss automaticity and its role in first language (L1) learning, including research into the lexical and semantic integration of novel words. Differences between second language (L2) and L1 learning are considered, especially with relation to the Complementary Learning Systems account. Two theories of L2 acquisition are then outlined and compared- the Revised Hierarchical Model and the Bilingual Integration Activation model. After this, L2 learning in the UK, and potential reasons for its poor performance, are proposed. My discussion then focuses on extralinguistic cues as an aid to language learning, starting with research on the use of pictures and gestures in L2 learning. I will then shift my attention to the general uses and benefits that gestures can have across many aspects of cognition, in particular speech integration. Following this, the use of iconic gestures in aiding the integration of newly learnt L2 words are discussed, alongside potential explanations for this gestural advantage, including Dual Coding Theory, Motor Trace theory, self-involvement, disambiguation and privileged access. I finish by providing a summary of my research aims.

Language learning

There are many aspects that must be mastered to become fluent in a language. It is not only the semantics of words that must be learnt, but the phonology (the systems and patterns of speech sounds), orthography (the written representations of language), morphology (the internal form and structure of words; Bahr et al., 2020) and syntax (the structure of words and phrases to create sentences). Despite the large range of areas that must be attained, many of these can be acquired relatively quickly (Barbir et al., 2023; Goldstein & Schwade, 2008; Wang et al., 2003). Furthermore, once acquired these aspects of one's first language appear to be processed automatically. In the following section, I will introduce the term automaticity before discussing this concept in relation to language processing and learning.

Automaticity

Automatic processing has been defined as cognition that does not require conscious effort or control from the subject and does not demand attentional capacity. In line with the latter, Newell (1990, p.136), described an automatic process as being "unstoppable (ballistic) and independent of the amount of information being processed".

Additionally, automatic processes are efficient, fast-acting and do not use limited-capacity resources (Hartsuiker & Moors, 2017; Holle & Gunter, 2007; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Schneider and Shiffrin's (1977) two-process theory of information processing describes automatic processing occurring without intention or awareness. However, it is important not to equate automaticity with fast non-automatic processing; "Rather, automaticity refers to a significant change in the way processing is carried out (some form of restructuring)" (Segalowitz, 2008, p.403). Automaticity has been identified across a number of different cognitive processes including aspects of reading (Samuels & Flor, 1997), the competition of attention (Stroop, 1935) and activation of lexical memory (Anderson & Bower, 2014; Conrad, 1974; Warren, 1972, 1974).

Early research into automaticity viewed an automatic process with an all-or-nothing approach. A process was either, rapid, uncontrolled with no attentional effort or it was voluntary, slow and required effort. Later work challenged this approach to automaticity, with Cohen et al. (1990) suggesting that processes can have varying degrees of automaticity, called the strength of processing (Garrod & Pickering, 2007). Another limitation with the all-or-nothing view, as argued by Bargh (1994), is that when studying complex cognitive processes, there are multiple components that work together. Bargh (1994) proposed that some of these components may be automatic whereas others may be more controlled. One such process, that comprises of multiple cognitive components, is language processing

Automaticity in language processing

Many components of language processing appear largely automatic. Without conscious effort, native speakers automatically process and understand spoken language and it is very difficult to actively try to avoid this (Hartsuiker & Moors, 2017). The process of speaking and listening is well practiced and a key function in our daily lives. Additionally, the process is quick, with a normal speech rate for adults being 150 words per minute (Maclay & Osgood, 1959, as cited in Levelt, 1989) and with a very low error rate of 1 per 1000 words (Garnham et al., 1982, as cited in Levelt, 1989).

The different levels of language processing, comprehension and production, can be investigated more closely to identify what components are and are not automatic. Language production consists of the conceptualisation of what will be said, the formulation on both a syntactic and lexical level and then construction on a phonological level (Hartsuiker & Moors, 2017). This is also in line with Levelt's (1989) framework of production which identifies three main stages; conceptualisation, formulation and articulation. Some aspects in language production are clearly controlled processes that are not automatic, such as constructing what to say (conceptualisation). But other aspects, such as lexical access and phonological encoding are viewed as having automatic properties (Hartsuiker & Moors, 2017).

Language comprehension begins with the interpretation of the input language, before processing understanding at a syntactical, lexical and phonological level. Many aspects of comprehension are considered to be automatic, given its fast and effortless nature. However, it has been argued that this process is not entirely automatic given that, in some cases, comprehension can be affected by structural ambiguity and high complexity of sentences (Hartsuiker & Moors, 2017). Overall, some aspects of language processing are automatic, such as the comprehension of simple speech and phonological encoding in speech production, and other components, such as checking speech for grammatical errors and determining subtly conveyed messages, are not automatic and do require conscious effort (Hartsuiker & Moors, 2017).

Automaticity in language learning

The following section explores how and when newly learnt words start to become automatically processed. According to the complementary learning systems (CLS) model of language learning, there are two systems that are involved in this learning process: one contributes to rapid integration of words and sentences and the other aids the formation of deeper meanings, generalisable connections and longer retention (Davis & Gaskell, 2009; McClelland et al., 1995). The CLS account states that new words are encoded immediately into episodic memory, which is mediated through the highly malleable hippocampal system. This hippocampal memory system also involves pattern-separated representations to differentiate between similar sounding phonemes and words. Additionally, novel items are encoded in a contextualised, episodic manner (O'Reilly et al., 2014).

Over time and repeated exposure, there is a shift to the neocortical system, with semantic connections being formed for the novel words. These semantic representations are generalisable and go beyond the initial contexts in which the words were learnt. Learning through this system is gradual and aids long-term retention in the cortical semantic store. The

CLS proposes that an offline consolidation period is crucial for these neocortical representations of novel words to develop and to integrate into the mental lexicon. The mental lexicon refers to the store of all known words in a person's vocabulary that they have available to them. Once novel words have been integrated into the mental lexicon, they can begin to interact with existing words by activating them or creating competition when being retrieved. These are automatic processes as they occur rapidly, unintentionally and without conscious effort.

The two systems proposed by the CLS (the hippocampal system and the neocortical system) work together to achieve language proficiency. Based on the CLS, words start to display automatic processing, such as lexical competition, once the shift from the hippocampal system to the neocortical system has occurred, following a period of offline consolidation. The next section turns to research measuring the emergence of these automatic processes following the learning of novel words in one's first language.

Novel word learning

In this next section, I will discuss novel word learning in adults and assess when lexical and semantic integration appear to take place (immediately after learning and/or after a consolidation period) and what methods are used to measure these automatic processes.

Novel word learning research is most often conducted with adults and involves a process of teaching and testing recall for new 'legal' first language words. These words are created independently from existing words and are designed to be pronounceable, to conform to language rules and to not evoke semantic associations with other real words (Deacon et al., 2004). This type of research does pose some differences to second language learning in that the phonology, orthography, morphology and syntax are not novel. However, research into novel L1 learning can give us insight into how novel words are integrated into an adult's pre-existing mental lexicon and semantic networks by measuring the presence of automatic processes (e.g., through lexical competition and semantic priming tasks). This research on novel L1 learning can often be applied to second language learning.

Additionally, unlike second language learning studies, a large area of research into novel word learning is not concerned with teaching the semantics of these words. Rather, the aim is often to determine how novel word forms are integrated into the mental lexicon, how long this integration takes, how to measure this integration and how these new words interact and often interfere with pre-existing words. Research in this area states that the lexicalisation of novel words can be tested by their word-like behaviours. One of these key behaviours is their ability to create lexical competition. Such ability (activation and competition) is believed to demonstrate the shift from episodic representations to lexical ones.

Lexical integration

Much research has been produced looking into the lexicalisation and consolidation of novel words (for review, see Palma & Titone, 2021). A number of different procedures have been used in the research in order to assess the lexicalisation of novel words. One of the most common is a lexical decision task in which participants have to respond to stimuli (identify a certain feature about the word) as quickly and accurately as possible. This sort of task is based on the idea that if new lexically neighbouring non-words have been successfully lexicalised, there will be a shift in the uniqueness point of the neighbouring word (that the non-word has derived from). The uniqueness point of the known word will shift to later in the word's pronunciation stream to account for this new lexical competitor. This newly learnt competitor would therefore delay recognition of the known word, known as lexical inhibition or lexical competition.

Being one of the earliest studies in this area, Gaskell and Dumay (2003) carried out a set of experiments testing how newly learnt words can create lexical competition. First language base words (e.g., *cathedral*) and nonwords were presented together in an initial phoneme monitoring task (familiarisation phase). The nonwords either deviated from real words at the final vowel of the word (final-deviation condition, e.g., *cathedruke*), by the onset of the word (initial-deviation condition, e.g., *yothedral*) or were not related to a real word at all (control). Reaction times in a lexical decision task showed that inhibitory effects for the final-deviation word items began only 3 days after initial exposure.

In their final experiment, lexical inhibition was tested after exposure using a different paradigm, pause detection. A silent pause was digitally inserted into the real base words and filler words. Additionally, time for lexicalisation was assessed as this experiment involved testing after exposure, and then again, a week later. Reaction times to indicate if a pause was present or not were recorded and found no differences between the control and final deviation words for the day 1 task. Lexicalisation effects emerged 1 week later, with the novel final deviation words having longer reaction times than the control trials. This suggests that, even without any further exposure to the novel items, these new words can be integrated into the mental lexicon after an extended period of time after initial exposure.

Overall, Gaskell and Dumay's (2003) research suggested that these lexical competition effects only emerge after several offline consolidation periods as they found no immediate lexicalisation effects. This supports the claim from the CLS that for these automatic integration effects to be evident, a period of offline consolidation is necessary. This research is also in line with a number of other studies that have also shown the importance of such a consolidation period for lexical competition (Dumay & Gaskell, 2007; Tamminen et al., 2010). Additionally, research has demonstrated the longevity of lexicalisation with lexical competition effects still being present 8 months after initial exposure, with little additional training (Tamminen & Gaskell, 2008).

The role of sleep more specifically during an offline consolidation period for lexicalisation has also been investigated. In Dumay and Gaskell's (2007) study, participants learnt novel words either in the morning (am group) or the evening (pm group). The pm group therefore had a night's sleep before being retested 12 hours later and displayed reliable lexical competition effects in a speeded pause detection task. In comparison, the am group (who had 12 hours of wakefulness) did not display such effects. After 24 hours, when the am group had now experienced a night of sleep, both groups showed equal lexical competition effects. The benefits to learning were also demonstrated in an explicit free recall task, with scores improving after sleep for both groups.

However, research has also found that a sleep consolidation period is not necessary for lexical integration. Using a lexical decision task and testing at different time periods throughout one day (2.5 hour intervals), Lindsay and Gaskell (2013) found lexical competition effects within a single day. These competition effects were found during the second lexical decision task that was completed 5 hours after the initial exposure to the novel words, suggesting that fairly fast integration can occur. Lindsay and Gaskell (2013) suggest two possible explanations for why they found such effects without sleep consolidation. Firstly, they suggest that they used enhanced training on the novel words with retrieval practice compared to standard learning received in Gaskell and Dumay's (2003) study. Additionally, they used repeated exposure and testing on the neighbouring words in their lexical decision task. This suggests that the intensity of learning and the use of known first language words in assessment play a role in how quickly novel words can be lexically integrated.

Another method that has been used to investigate when lexical activation and competition emerges is the visual world paradigm (VWP). This involves tracking participants' eye movements as they listen to individual sounds, words or sentences and inspect either things in the real world or visual information on a computer display (Guan et al., 2019). Kapnoula et al. (2015) taught participants half of a set of novel non-words without meaning, and then measured the integration of these new words using the VWP. This involved creating three different auditory stimuli of a target word by splicing three similar words (e.g., using the target word 'job', a non-word 'jod' and a target word competitor, 'jog'). The audio recordings were then played alongside the presentation of four pictures. The participants' task was to click the corresponding picture to the target word- the fixation to which was taken as an estimate of its lexical activation. Longer fixation times for newly learnt non-word stimuli were found, indicating competition effects with existing words immediately after learning. The taught non-word condition also showed larger interference effects than an untaught non-word splice condition.

Magnuson et al. (2003) had similar findings using a VWP. By creating an artificial lexicon, the frequency of exposure to the novel words was able to be controlled and the effects on lexical competition were measured. Significant frequency effects were found, with higher frequency novel words suffering less lexical competition than those in the low frequency condition on the same day as learning. Further research has suggested that the level of exposure to the novel words could impact how effective sleep consolidation can be (Walker et al., 2019).

The research shows that evidence of rapid lexical integration of novel word forms into the mental lexicon can be found using certain techniques. It is clear that many factors, such as the number and intensity of learning sessions, frequency of exposure to the novel words and outcome measure used, have an impact on the speed and level of such integration and competition effects found. Overall, this process is still quick with studies finding evidence of integration from immediately after learning. This integration is robust and long-lasting with lexical competition still evident 1 week after learning with no additional exposure (Gaskell & Dumay, 2003) and 8 months with just 2 additional exposure sessions (Tamminen & Gaskell, 2008).

Semantic integration

In the real world, words are encoded by more than just their lexical properties. A vital part of word learning also involves understanding the associated meanings that accompany these new words. Fortunately, not all research has excluded the semantic aspect of novel word learning in their studies. Further research can tell us more about the process of learning a word's semantics and integrating it into existing semantic networks. In the following section, I will discuss the literature on the semantic integration of novel words. Through this exploration of the literature, a number of different paradigms and methods will be introduced that have been used to measure semantic learning, including explicit and implicit tasks. Additionally, this section will explore when such integration becomes evident and whether a consolidation period is necessary for this.

A range of methods are used in the word learning literature to determine if the semantics of novel words have been integrated. Often an explicit task is used to assess initial knowledge, these come in the form of a free recall task, cued recall or alternative forced choice task (AFC). Many studies use this very explicit, controlled retrieval measure alongside other more automatic tests. Such automatic tasks can allow us to see if word meaning can only be explicitly retrieved, or if it is integrated into semantic networks such that it can engage automatic processes with word-like properties. Within this thesis I will refer to learning displayed through such automatic-like processing (that demonstrates a shift from merely episodic memory that could be evident through slow, controlled revival) as deep semantic learning.

One such task is semantic priming (see McNamara (2005) for an overview), which occurs when a target word is preceded by a semantically related word, causing pre-activation of the target words semantic network. This is based on the concept that when the semantics of a word are activated, this activation is not confined to the single word node, but that it spreads to other similar concepts that share related semantic content. This pre-activation from the related prime word means that the target word is activated and recognised quicker than if the prime word was unrelated and thus not activating related semantic networks. This spread of activation can result from semantic association or semantic relatedness. Semantic priming paradigms are often used in language learning research as it can demonstrate if the meaning of a newly learnt word has been semantically integrated. Evidence for priming is provided when learnt words can elicit a priming effect (faster reaction times compared to an unrelated baseline) on related known target words.

Tamminen and Gaskell (2013) carried out both unmasked and masked semantic priming studies. By including the latter, it ensures that the semantic priming task is a test of the words integration into the semantic memory store as opposed to simply existing in the episodic memory. In these semantic priming experiments, the newly learnt novel words were used as primes in a lexical decision task using familiar words. Unlearnt non-words were also used as control target words. In both the unmasked and masked experiments, a priming effect of the learnt novel words was found, with this effect being stronger 1 week after learning than immediately after. This suggests that consolidation plays a key role in this semantic priming effect with the consolidation period allowing further development of semantic networks. These results are consistent with other semantic priming studies. Van Der Ven et al. (2015) found that novel words only primed semantically related words after 24-hours and did not find any priming effects immediately after learning.

As well as behavioural measures, neuroscientific techniques have also been used whilst performing priming tasks to measure the semantic integration of novel words. Bakker et al. (2015) recorded the electrophysiological responses to newly learnt novel words whilst performing a priming task. Lexicalisation was tested by the differences in N400 amplitude between the trained novel and existing words. These differences were found to significantly reduce after a 24-hour consolidation period, demonstrating that responses to the novel stimuli became more word like. The semantic priming effects found were more immediate, with semantically related primes producing greater Late Positive Component (LPC) responses than semantically unrelated primes in the tests before and after the consolidation period. This research suggests that consolidation is important for the lexicalisation of newly learnt words but that some semantic processing can begin to occur earlier and without the need for this consolidation period. This highlights the different levels of processing that are required in novel word learning.

A long with semantic priming, techniques using eye-tracking have also been used as a method for assessing the learning of novel words through picture-word associations. Weighall et al. (2017) taught both adults and children novel words using pictures of novel objects. A visual world eye tracking task was used to assess competition effects, i.e., fixations on novel learnt objects (e.g., ‘biscal’) when hearing existing words (e.g., ‘click on the biscuit’). The eye-tracking data found that for both adults and children, novel trained competitors were fixated at higher rates than controls or untrained objects. This competition effect was still smaller than that found for already existing competitor words, for both novel words learnt the previous day and those learnt immediately before test. These results suggest that competition effects can develop quickly but don't reach the same levels as established words. The effects of sleep consolidation appear less vital for this particular assessment of semantic integration.

Another method that has been used to measure semantic integration is a Stroop-like paradigm. In Geukes et al.'s (2015) study, participants were taught novel colour words through statistical association (statistical learning principle) and were then tested immediately after learning as well as the following day. A significant congruency effect was found on both days for the novel words, with faster reaction times for the congruent ink and word trials than the incongruent trials. This suggests that the semantics of the novel words have been integrated as they are causing a conflict of attention. In another experiment from this study, the native words were tested in isolation in the Stroop assessment, without the first language colour words. Additionally, a second group was added that completed both Stroop tasks on the second day to differentiate between practice effects or effects of consolidation. The congruency effect was no longer present in the Stroop task immediately after learning, and only appeared after the 24-hour consolidation period. In the second group that completed both Stroop tasks on day 2, both showed congruency effects, indicating that the effects found must be a result of difference in the passage of time and not practice effects. These results show that the congruency effect emerged quicker when related native words were also present.

Geukes et al. (2015) suggest that the immediate effects found in the initial study were a result of the additional exposure to the native words, which provided further learning context. Like much of the previous literature mentioned, they emphasise the importance of a memory consolidation period. Using a Stroop paradigm in such a way does mean that this research has a strong orthographic and phonological element to the assessment measure, but it still demonstrates conceptual learning even in a shallow learning environment with limited semantic context (only four colours used in each test block).

Similar to varying font colour, varying font size has also been used within a task to measure when automatic access to newly learnt L2 word meanings becomes available depending on the consolidation period. In Tham and colleagues' (2015) study, participants were taught pairs of Mandarin characters and L1 English animal names, either in the morning or evening. In the test session 12 hours later (either after a night's rest or a day of wakefulness) participants completed two tasks as a measure of integration. One, which presented two novel learnt Mandarin characters together, assessed size congruity effects across semantic size (the size of the animals) and physical font size. Quicker reaction times were recorded when the two characters were congruent across the semantic and physical dimensions in the evening learning group, showing integration of novel-form meaning mappings. However, these size congruity effects were not evident in the morning learning group who had not slept, demonstrating the importance of sleep consolidation for automatic semantic integration.

Overall, evidence of automaticity at the lexical and semantic level can occur soon after learning, potentially even immediately, under certain learning conditions. These include high levels of novel word exposure and frequency of the learning sessions. Another factor that can influence these immediate integration effects are the conditions of the assessment measure, in particular, if any additional learning context can be gained from the assessment. However, an offline consolidation period does appear to be beneficial for the integration of both lexical and semantic information, as evidenced in the meta-analysis by Schimke et al.(2021) which indicated that a period of sleep between learning and test produces greater learning effects than the same period of wakefulness. . Research has also alluded to the effects of the time period between learning and sleep on the impact of such a consolidation period (particularly in long-term explicit memory; Walker et al., 2019). Overall, an offline consolidation period is not *necessary* to find lexical and semantic integration effects given the right conditions and characteristics of learning and assessment, but it certainly does appear beneficial.

Second language learning

Before discussing automaticity and the role of consolidation in L2 word learning, it is important to first note that L2 learning differs from L1 learning in several regards. Standard L1 learning (without the presence of another language i.e. not bilingual or multilingual learning) takes place in a fully immersive environment, a key characteristic for language development. This full immersion means that the first language has ample time to be practised and for mistakes to be made. In contrast, in standard L2 learning, there is rarely substantial immersion, with learners typically having only limited exposure to the L2. Research has found immersion to be an invaluable tool in L2 learning, with even partial L2 immersion programmes leading to greater acquisition (Bergström et al., 2016). The benefit of immersion is not only evident in vocabulary acquisition, but in other areas of language learning too.

Freed et al. (2004) investigated the different learning contexts for improving oral fluency in L2 learners and found that both an intense immersion programme and a study abroad programme outperformed the use of formal language classroom sessions in a home institution. Interestingly, students on the immersive course had greater gains in oral fluency than students who took the summer abroad programme in the L2 country. One explanation given for this is that the immersive-course students reported using the L2 language significantly more in out-of-classroom activities than the summer abroad students. This research suggests that it is the amount of time spent exposed to, and using, the L2 that is particularly beneficial when in an immersive language setting. Additionally, the sessions delivered in the immersion programme were in a shorter time frame (7-weeks) than for the summer abroad programme (12-weeks) and this difference in distribution could also have an impact. Overall, it is evident that an intense, immersive learning experience involving little use and exposure to the L1 is beneficial for L2 development.

Another challenge faced when learning an L2 is having to learn a whole new set of phonemes and orthography. Although these must be learnt during first language acquisition as well, there is no conflict with pre-existing knowledge. With regards to learning new orthography in L2 acquisition, this can be very challenging, as languages take many different forms. For example, the L2 being learnt may use an entirely different writing system and alphabet to the existing language. Hayes-Harb and Barrios (2021) reviewed the large body of literature on this topic and highlighted the interference effects found between L1 and L2, particularly with regards to orthography and phonology. The authors outlined a number of variables (e.g., the familiarity of L2 graphemes) that appear to have the greatest influence on L2 phonological development. Similarly, sentence structure with regards to word order is another area where differences are often experienced in second language learning. For example, languages such as English, French and Mandarin typically use a subject-verb-object word order, whereas Latin, Japanese and Hindi use subject-object-verb and Arabic, Hebrew and Welsh use a verb-subject-object structure (Al-khresheh, 2010). Interlingual interference can therefore occur when individuals transfer L1 structure onto L2 sentences, thus causing word order errors and creating additional hurdles (Al-khresheh, 2010). Additionally, languages often have collocational differences (the set order of words to create a common phrase, e.g., in English we say, ‘to *make* a mistake’, not ‘to *do* a mistake’). This adds an additional level of difficulty to mastering a second language as L2 learners will often incorrectly apply L1 linguistic knowledge to an L2 context (Sadeghi, 2009).

Consolidation and processing in L2

In addition to these challenges, research has also suggested that L1 and L2 differ on a processing level. Palma and Titone's (2021) review included a section on L2 lexicalisation which discussed the negative impact that multiple languages can have on language representation and processing. With regards to the latter, reduced lexical access and quality during L2 processing have been proposed compared to the more prominent L1 (Palma & Titone, 2021). However, the amount of interference of this impact from the L2 is of course very dependent on proficiency (Lindsay & Gaskell, 2010). Applying the CLS to L2 learning, it would be unlikely for L2 words that have been learnt on the same day as testing to display any interference with the L1 due to the words being mediated via the hippocampal system.

The CLS account for L2 word learning suggests that the access to new L2 representations through the hippocampal system will be much slower than for existing L1 representations that have already been consolidated neocortically (Lindsay & Gaskell, 2010). As in L1 representations, L2 lexicalisation is thought to benefit from periods of sleep which aid offline consolidation, a claim supported by research that has found sleep to improve L2 vocabulary learning (Gais et al., 2006) and L2 semantic form-mapping (Tham et al., 2015). Following such consolidation, the development and strengthening of L2 neocortical representations can occur. In this way, and continuing to apply the domain-general CLS account to word learning, Lindsay and Gaskell (2010) state that once developed, the lexicons for L1 and L2 theoretically should not possess distinct differences, cognitively or neurally. However, the authors outlined a number of factors that could potentially contribute to differences. These include age of acquisition, proficiency, frequency of exposure to each of the languages and lexicophonological structure which can all influence the representations created.

Theories of L2 acquisition

Throughout the years of L2 learning research, many theories have been developed on the bilingual lexicon (see Jiang (2015) for an overview). These theories typically have a focus on a certain aspect of bilingualism including learning, processing, storing, representations and accessing a second language. Accordingly, in the following section I will be focusing on the Revised Hierarchical Model (RHM) as it is specifically a model for translation word production and so is relevant to my research. Additionally, unlike other models which may focus more on advanced bilinguals, the RHM considers the differences between low and high proficiency L2 learners. I will also review the scrutiny the model has come under, including language selective access and separate lexicons, the mediation of translation from L2 and evidence of strong conceptual connections from L2. Finally, I will discuss the proposed alternative model, the Bilingual Interactive Activation (BIA) and the updated model (BIA+).

Revised Hierarchical Model

Kroll and Stewart's (1994) Revised Hierarchical Model (RHM) of bilingual memory combines theories of word association and concept mediation (as outlined by Potter et al., 1984). The model outlines separate lexicon storage for first and second language and describes the stage-like process of L2 learning with the development of connections to the L2 lexicon store. The RHM predicts key differences between early (low proficiency levels) and later L2 learners (higher proficiency levels) in terms of these connections and translation production.

The RHM proposes that the L1 has privileged access to conceptual meanings. The L2, on the other hand, requires lexical mediation through the L1 translation until the L2 becomes developed enough to access conceptual links directly (Sunderman & Kroll, 2006). As a result, initially, strong lexical links are created from the L2 store to the L1. With conceptual links between the L2 and meanings being weak during early learning, the strong lexical links must be heavily relied on during translation from L2 to L1 (backwards translation). Therefore, in early-stage learners, knowledge about the L2 is mediated via the strong lexical links created with the L1, which then access the knowledge through the strong preexisting conceptual links with the conceptual information itself (Figure 1, left). This is termed ‘word association’, as the salient form of interconnection between the two languages is lexical (Kroll & Curley, 1988). After repeated exposure to the L2, the direct conceptual links between L2 and conceptual knowledge strengthen and eventually, once the learner has gained greater proficiency, this route to knowledge of concepts is taken automatically (Figure 1, right).

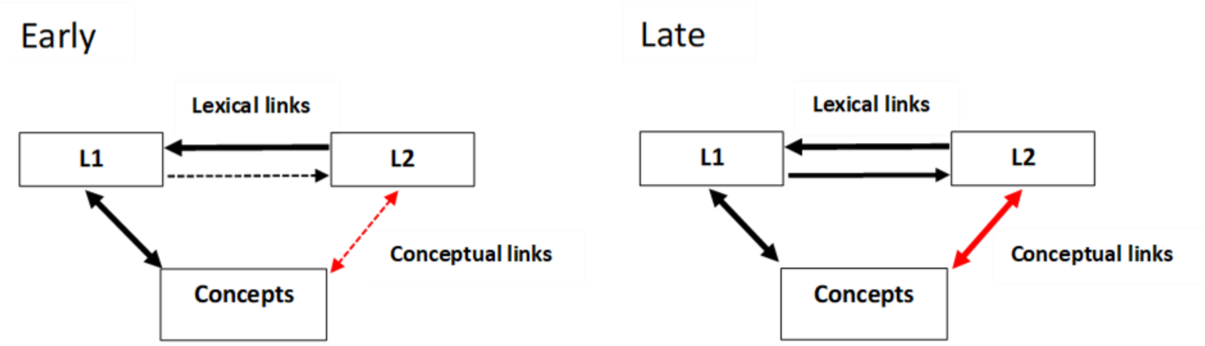


Figure 1. Hierarchical revised model of lexical and conceptual representations in bilingual memory (Kroll & Stewart, 1994), separately for the early learning stage (left) and proficient L2 speakers. Thin dashed lines represent weak links.

Criticisms of the RHM

Due to conflicting research across several empirical papers, including that from Kroll and Stewart (1994) themselves, the RHM has come under criticism. Brysbaert and Duyck (2010) argue that the RHM is no longer the most accurate way to categorise the processing of concepts across two languages. The paper cites a number of limitations of the model and claims that it cannot account for recent empirical research that has prevailed in the 15 years since the model's publication. In the following subsections, I will briefly outline some of the key criticisms along with the proposed alternative model, the BIA, as well as summarising the response given by Kroll and colleagues (2010) to address these critiques.

Language selective access and separate lexicons

Firstly, the RHM does not assume lexical nonselectivity (that words from different languages are stored in a single integrated mental lexicon- allowing activation across languages). The model incorporates separate lexicons for different languages. However, this assumption has been criticised by Brysbaert and Duyck (2010) who argue that the model cannot account for the recent evidence for nonselective lexical access. At the time of the RHMs publishing, there was little research into language nonselectivity, and the handful of empirical studies that existed gave varied conclusions. Since then, research has emerged that suggests parallel access between two languages in many aspects of language processing including visual word recognition (Dijkstra, 2005).

However, as Kroll and colleagues (2010) point out, evidence of parallel access does not necessarily equate to evidence of an integrated lexicon. There could still be two lexicons that "are functionally separate but with parallel access and sublexical activation that creates resonance among shared lexical features" (p.2). Kroll and colleagues (2010) add that that parallel activation may be a general feature of language access as it can be observed in bilinguals for whom their two languages do not share written scripts and therefore lexical items cannot be treated as if they are the same language (Emmorey et al., 2008; Hoshino & Kroll, 2008). Additionally, since the emergence of greater evidence for nonselectivity of lexical access after the publication of the RHM, this new evidence for nonselectivity was addressed by Kroll and De Groot (1997) by incorporating parallel activation into the RHM.

Translating from L2 to L1

Another argument against the RHM comes from research that has questioned how backwards translations (from L2 to L1) are mediated. The strongest evidence against the RHM would be data that counteracts the model's prediction that, for L2 learners with weak proficiency (early-stage learners), only translation from L1 to L2 would reliably be conceptually mediated and not backwards translations. The model assumes that conceptual links from the L2 words develop with greater proficiency (later in learning) and that initially, backwards translations are mediated by lexical links to the L1. According to the model, at this stage of learning, only forward translation should be capable of consistently relying on conceptual links.

However, Brysbaert and Duyck (2010) argue that there is evidence that shows both forward and backwards translation production can be affected by several semantic properties (see Kroll & Tokowicz (2005) for overview). Research has demonstrated sensitivity to semantic effects in both directions and therefore translation production must be, at least partially, conceptually mediated. One such study by De Groot et al. (1994) assessed forward and backward translations in bilinguals and found that both directions were affected by meaning variables.

Brysbaert and Duyck (2010) state that these results challenge the RHM as they describe the data as coming from bilinguals with "limited fluency in their second language" (p.366). However, the undergraduate students used in this research were described in De Groot and colleagues' (1994) paper as being "all relatively fluent in their second language" (p.608) with their university course including aspects of English. Additionally, the participants in this study gave measures of self-assessed proficiency levels before the experiment on a scale of 1-7 (with 1 being very low and 7 being equivalent to their native language). Across the different forms of translations (forwards and backwards), the bilinguals in this study averaged a rating of just below 5 for comprehension and production abilities.

As Kroll and colleagues (2010) argue, these participants, as well as many other Dutch-English bilinguals that have contributed to a large body of the research in this area, were highly proficient bilinguals who had vast exposure to their L2 through media and also completed university education in both languages. Based on the RHM, one would expect individuals with this level of proficiency to have developed stronger conceptual links and thus conceptual mediation for translations would be supported in both directions.

There is evidently still debate around the level of proficiency of this demographic group. Within the literature this 'developmental shift' from L2 lexical mediation to completing translations through conceptual mediation has been discussed and yet it is still difficult to definitively pinpoint. Kroll's earlier work (Kroll & Stewart, 1994) mentioned a critical point during learning for this developmental shift and how this change depends on many factors, including how many years the L2 had been studied for. Kroll's work has highlighted that the RHM is of course limited to late bilinguals that have created a distinct L1 lexicon before the introduction of an L2.

A further criticism for the RHM however is that De Groot and colleagues' (1994) study appears to show equivalent conceptual mediation across both directions. They concluded that meanings played a slightly larger role in forward translations than backwards but that, overall, there appeared minimal differences between the two directions. This raises issues for the RHM, as only highly proficient, balanced bilinguals are proposed to have such equivalence of conceptual mediation (Sunderman & Kroll, 2006).

However, it has been questioned whether even these highly proficient bilinguals can ever meet native like conceptual connections from the L2 (Kotz & Elston-Güttler, 2004; Silverberg & Samuel, 2004). Alongside their model, Kroll and Stewart (1994) also included research with fluent Dutch-English bilinguals in forward and backward translation tasks. The results are somewhat conflicting to the model, with differential semantic effects on translations in both directions in fluent Dutch-English bilinguals. One explanation given by Kroll at al. (2010) is that the items used in the 1994 experiment were lower frequency than items used in other studies. They argue that some effects may be the result of ease of item processing, which can simulate language skill effects (Kroll & Tokowicz, 2005). Low frequency items or complex words such as abstract noncognates can all lead to asymmetry in the processing of forward and backward translations. As a result, even highly skilled bilingual individuals may display inconsistency in the language processing under certain conditions depending on the words used and the context of the task.

Strength of connections between L2 and concepts

Following on from this, another similar issue raised by Brysbaert and Duyck (2010) is that the connections between L2 and concepts are potentially stronger than the RHM proposed. Since the publication of the RHM, research has emerged that has found that less proficient L2 learners can gain access to meanings directly and show semantic sensitivity (Dufour & Kroll, 1995; Sunderman & Kroll, 2006), the latter will be explored in detail in the section 'Comparison between RHM and BIA' (p.21). This evidence is obviously problematic for the model as the RHM assumed that the conceptual links between the L2 and meanings were weak in low proficiency learners. This weak link was initially believed to be bidirectional (see Figure 2, left). However, although the research from Sunderman and Kroll (2006) shows that low proficiency bilinguals may be able to access conceptual information directly, the learners were not able to name these concepts in translation production tasks. This important distinction between production and recognition is one that Brysbaert and Duyck (2010) failed to make.

Kroll and colleagues (2010) stand by the RHM's original stance that this link between L2 and concepts is initially weak. But they account for the subsequent research by acknowledging that the model incorrectly assumed the weak link to be bidirectional. They determine that the weak link is asymmetric in the sense that a low proficiency learner may find L2-to-concept tasks easier than concept-to-L2 that require translation production (Figure 2, right).

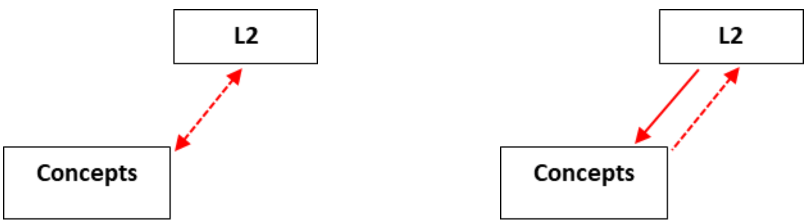


Figure 2. Diagram demonstrating the previous bidirectional weak link between L2 and concepts proposed by the RHM (left) and the possible asymmetrical weak link between the two (right) as described in Kroll et al. (2010). Thin dashed lines represent effortful connections, thin solid lines represent slightly less effortful connections. Note that the lines are still thin, indicating the weak link.

The BIA/BIA+ model

Brysbaert and Duyck (2010) argue that the RHM cannot be adapted to overcome the criticisms they discuss and argue that the more productive approach to finding an accurate model for bilingual language processing is to base it upon existing, well-established models of monolingual language processing. With that in mind, they suggest that another model, the Bilingual Interactive Activation model (BIA), is a more accurate model that can account for the recent research. The BIA, which was first introduced by Dijkstra and Van Heuven in 1998, was based largely on the Interactive Activation model for monolinguals (McClelland & Rumelhart, 1981). Initially, it was a model for bilingual visual word recognition, but it is also thought to be applicable to other domains such as spoken word recognition.

Unlike the RHM, this model is non-selective, with a single integrated lexicon store. The model proposes that nodes at multiple processing levels are either activated or inhibited during recognition (see Figure 3). When a word (or nonword) is presented, the features of the letters (e.g., where they are positioned within the word) activate letter nodes that match these features whilst also inhibiting letters that do not match the features. These letter nodes then excite words that contain these letters in the specific position. Due to the model being non-selective, words in both languages can become excited if they contain the same letters in the same position, which can create lexical competition between languages. Other words that do not contain the correct letter positions are inhibited. While all words inhibit each other (irrespective of language), activated words feedback activation to the letter nodes. Finally, the word nodes activate the language nodes for their corresponding language and these language nodes inhibit words from the other language.

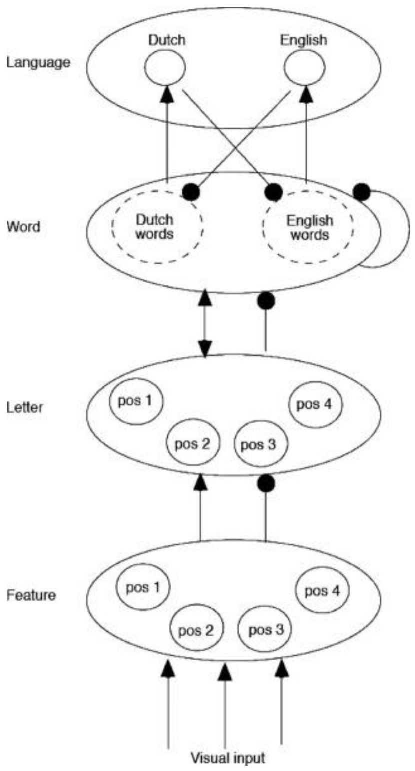


Figure 3. The BIA model.

This model also differs from the RHM in that the focus is on how distinctions are made between languages, when reading or listening, in the advanced stage of second language learning - although Dijkstra and Van Heuven (1998) have suggested that the system can be affected by varying levels of proficiency. They highlighted that the time it takes to complete this word recognition is depended on many characteristics of the word, including its frequency of exposure. This suggests the BIA model does consider the different processing of varying levels of proficiency, but that it does not outline these differences explicitly. It is suggested that with lower levels of proficiency, less inhibition and activation would take place.

The model can account for a large body of research in the area of bilingual word recognition (for an overview see Dijkstra & Van Heuven, 2002) (Dijkstra et al., 1998). However, the BIA model does have its shortcomings as it fails to incorporate semantic and phonological information into the model. Additionally, the effect of task demands during word recognition are not made clear. To account for the emerging research that posed questions to the BIA, Dijkstra and Van Heuven (2002) created an updated model, the BIA+. This model incorporated the previous BIA, but with additions to include phonological and semantic information and processing. Like with word processing for the BIA model, a similar process is taken in the BIA+, with multiple orthographic word candidates being initially activated. These activated representations then excite the associated phonological and semantic representations. Due to the frequency of exposure, L2 phonological and semantic representations are believed to be delayed in activation compared to the L1, known as 'temporal delay assumption' (Dijkstra & Van Heuven, 2002). All three representations influence the language nodes that become activated.

Another addition is that the BIA+ has a separate, distinct system for word identification and task/decision schema. Unlike the word identification system, the task schema is non-linguistic specific and accounts for other factors that can affect the processing of a task involving word recognition. This could include "instruction, task demands or participant expectancies" (Dijkstra & Van Heuven, 2002, p.187). The decision schema regulates control and can adapt the decision criteria.

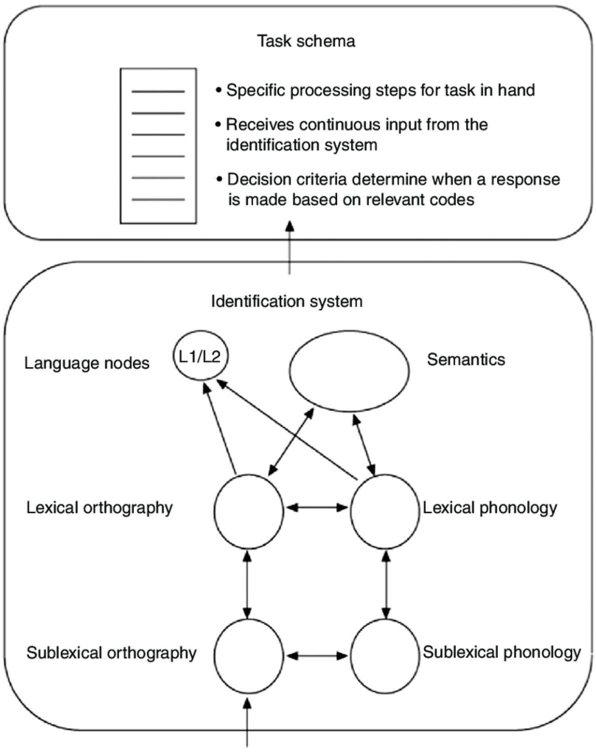


Figure 4. The BIA+ model.

Comparison between RHM and BIA

There is limited research that attempts to directly compare predictions across the two models. One that does, by Sunderman and Kroll (2006), focuses on answering some key differences between the models. One area where both models make different predictions is with respect to activation of phonological neighbours. The RHM predicts that in early learners, activation of an L2 word results in activation of the L1 translation and its word form neighbours, but not of the L2 word form neighbours. In contrast, the BIA would predict phonological co-activation in this case. Sunderman and Kroll (2006) aimed to investigate if both forms of interference occur (activation of lexical form relatives to the L2, as well as lexical form relatives to the L1 translation equivalent). They also investigated if the access of the concepts of L2 words increased with proficiency.

In the study, two groups were tested on a translation recognition task (are the L2-L1 word pairs translations or not). One of the groups contained participants with a lower Spanish (L2) proficiency compared with the higher proficiency levels of the other group. The task contained critical trials of foils with incorrect translation pairs that varied by form condition and meaning (see Table 1 for an overview of conditions). For example, for the correct translation pair of cara-face, the critical distractors included a form-related neighbour to the L2 word (e.g., cara-card), a form-related neighbour to the L1 word, the translation equivalent (e.g., cara-fact), or a meaning-related word (e.g., cara-head).

Table 1

Illustration of the different foil words used in each condition for the pair cara-face, adapted from Sunderman and Kroll, 2006.

	Form conditions		Meaning condition
Grammatical class	L2 neighbour	L1 neighbour	Semantic
Same	card	fact	head
Different	care	fast	pretty

One key finding was that even less fluent L2 learners experienced interference from the meaning-related pairs. This appears to contradict the RHM assumption that low proficiency learners have weak links to conceptual knowledge and therefore do not access conceptual mediation during translation. However, a possible explanation for these results in terms of the weak but asymmetrical link between L2 and concepts in line with the RHM, has been explained earlier (Criticisms of the RHM, p.18). Another interesting finding was that only less proficient L2 learners displayed interference effects of form relatedness through the L1 translation equivalent (e.g., cara-fact). This suggests that only these lower level learners were activating the L1 translation equivalents during the translation recognition task, thus supporting the claims of the RHM (Kroll et al., 2010).

Sunderman and Kroll (2006) also found results that supported the predictions of the BIA/BIA+ model, with all learners presenting inhibitory effects from form-related lexical neighbours. This form-related lexical interference was modulated by grammatical class, with the interference effects disappearing when different grammatical class was used across the two words. The inclusion of the task schema in the updated BIA+ model can account for these findings. Unrelated lexical form neighbours are unlikely to produce overlapping meaning activation and therefore information on grammatical class may be used as a cue to word identification. When grammatical class information does not match, this can be used by the task scheme as a cue to judge that the pair are not translation equivalents, thus the interference from lexical neighbours diminishes. Sunderman and Kroll (2006) summarise this by saying “By this explanation, the initial bottom-up activation of form neighbours is blind to grammatical class; only later in processing is that information used as a criterion for making a response” (p.416).

It is important to consider the fact that although the models discussed are both for the bilingual lexicon, they have different focusses and functions. The RHM is a model for word production, proposed for translation production, whereas the BIA/BIA+ model is for the process of visual word recognition. Despite being similar, with translation production tasks involving a combination of recognition and production, these two processes still contain many differences (see Kroll & Dijkstra (2002) for an overview of these similarities and differences). For example, in word recognition (in the form of visual comprehension) orthographic codes bear a larger and earlier role in the process than for production. In production, phonological codes are more critical. Additionally, the two models consider different stages of L2 proficiency. The RHM is a model of language development and describes how translation word production processes can alter with increased proficiency. Whereas the BIA/BIA+ model is an account of visual word recognition for proficient bilinguals with developed L2 lexicons.

In conclusion, much research has been conducted on the bilingual lexicon and although the RHM has faced criticism, it is still a profoundly influential model for bilingual word production with regards to translation performance. The proposed alternative model, the BIA+, is not a model for the same process. With regards to my research on the translation of newly learnt L2 words, the RHM is the most appropriate model.

Besides the afore-mentioned processes and challenges, there are a variety of other factors involved in the acquisition of a second language. A reoccurring factor that is mentioned throughout the literature as impacting acquisition and processing in L2, is the frequency of exposure. Exposure to L2 vocabulary is far greater in an immersive environment. However, such an environment is not always available to learners. This means that for some nations, including the UK, L2 learning has become a nationwide struggle. The difficulties of learning an L2 within the UK, as well as the importance and implications of this in today's world of global communication, will be discussed in the following section.

How the UK is falling behind

In the UK, only 32% of 16-30-year-olds are able to confidently read and write in another language, whereas the European average is 89% (Bowler, 2020). Research from the British Council has found similarly low figures, with 75% of adults unable to speak a sentence in one of the top 10 most economically important non-English languages (Tinsley & Board, 2017). A recent survey, also conducted by the British Council, found that just over a quarter of UK adults regret not learning another language fluently (Gough, 2023). When it comes to determining why the UK are so far behind the rest of Europe, there are several possible reasons.

The first is that in being an island, the UK is geographically separated from the rest of Europe. This means there is little exposure to other languages unless residents travel overseas. To add to this, English is a dominant language globally, with roughly 1.35 billion people speaking it worldwide (Eberhard et al., 2023). Therefore, when people do travel abroad from the UK, although they may be exposed to another language, they can often 'get by' with just speaking English, thus heightening the perception that other languages are not necessary. These factors limit exposure to other languages in the UK and therefore a heavier reliance is placed on the education system to learn second language skills.

A second key reason for the UK's short fallings is that foreign language learning is not mandatory in UK schools in the same way that it is in other European countries. Although the UK does have modern foreign language learning as a compulsory subject, this begins much later, and for far fewer years than in other countries (Delhaxhe, 2012). In most European countries, compulsory foreign language learning starts between the ages of 6 and 9 years old. A handful of countries, such as Belgium and Spain, even start at pre-primary education at the age of 3. However, in the UK, taking a modern foreign language subject does not become compulsory until 11 once students enter secondary school (Delhaxhe, 2012). Additionally, compulsory foreign language learning continues until the end of secondary school in all European countries, apart from Malta and the UK (Delhaxhe, 2012). Unlike the rest of Europe, which has been increasing the number of years of compulsory foreign language learning, the UK reduced the number of compulsory years. Compulsory foreign language learning was originally specified as 11-16 years, when the legislation was passed in 1988. However, since then, more flexibility in the curriculum for 14-16 olds years has been introduced (in 2004 in England), which has meant that after 14, the study of foreign languages is now optional (Rodeiro, 2009). This change in policy has led to less than half of UK students to continue to study a foreign language beyond the age of 14 (Goddard, 2018), compared to 98.6% of pupils in the EU (excluding the UK) learning a foreign language in secondary school in 2014, of which 60% were learning two foreign languages (Eurostat, 2016).

In an effort to reverse this trend of declining numbers and give students a well-rounded education of core academic subjects, the English Baccalaureate (EBacc) performance measure was introduced by the UK Government in 2010 (Long & Danechi, 2019). As the EBacc is a performance measure for schools, students are often encouraged by the school to enter for the EBacc by choosing a language subject and a humanities subject at GCSE as well as core English, Maths and Science. The number of students in state-funded schools who were entered for the EBacc was 39.3% for the academic year 2022-2023 (National Statistics, 2023) a figure that has remained relatively unchanged since 2013 (Long & Danechi, 2019).This is still short by some measure of the Department for Education's ambition that 90% of year 10 pupils in state-funded schools study for the EBacc by 2025 (Department of Education, 2022). Perhaps most importantly, the EBacc was never made compulsory for students and, despite initial plans, did not become an official qualification, meaning students gained no extra credit for achieving the EBacc.

The UK education system's poor commitment to second language learning is also demonstrated in the British Council's annual Language Trends surveys which are used to gauge modern foreign language uptake in schools and pupils' attitudes towards second language learning. The 2020 survey showed the paucity of second language learning taking place in UK primary schools, with only 30% of teachers stating that time allocated to second language learning was over 45 minutes a week. As well as this, almost 40% of primary schools admitted that students do not always receive the full allocation of second language learning lessons, with timetable restrictions, school trips and pressure on high performance in SATS often listed as reasons for this reduction in language teaching (Collen, 2020).

The most recent Language Trends survey, which was carried out in February and March 2023, included responses from teachers at 575 primary, 586 secondary and 155 independent schools in England. It continued to investigate the current situation of language teaching and learning in their schools (Collen, 2023). In a similar vein to the 2020 survey, one third of respondents reported that, over the school year, there had been a negative impact on time allocated to language learning because of pupils spending extra time on literacy and numeracy. Additionally, 71.8% responded saying that they had not received funding to develop resources for teaching modern foreign languages. Another issue with second language learning in the UK that was highlighted in the report is the lack of consistency in the language taught to pupils throughout their time in education. Only 3% of secondary schools reported that all pupils continued learning the same modern foreign language that they had been taught in primary school, demonstrating the lack of continuity in students' L2 learning.

The UK's recent withdrawal from the European Union (EU) at the beginning of 2020 may also have had an impact upon how necessary UK citizens feel it is to learn a second language. This appears to have affected foreign language learning in state schools to a greater degree than in independent schools. The 2020 Language Trends survey found that in state schools, 31% of teachers said that languages were less valued by the school community (this includes senior management teams, parents and governors). One state school teacher stated that "some parents are not supporting language learning because of Brexit" (Collen, 2020, p.16). Another question in the survey asked teachers if "the UK's recent departure from the EU had an impact on pupil motivation to study languages at your school?" It was found that 53% of state schools responded that pupils were less motivated to study European Languages. However, the same response was given from 28% of independent schools, with 70% of independent schools reporting no impact on pupil motivation.

Teachers at independent schools reported that hosting international pupils and utilising language assistants really benefited the teaching of foreign language subjects. However, such initiatives are less common in state schools- independent schools are twice as likely as state schools to take part in student exchanges and employ language assistants and are over three times more likely to have opportunities for work experience abroad (Collen, 2020). As reported in the Language Trends survey, state schools do not have the funds or ability to provide such immersion in foreign languages, with many students being "culturally deprived" and never speaking the foreign language outside of the classroom (Collen, 2020, p.16). The survey also revealed that 27% of state schoolteachers reported that funding for initiatives in languages had been decreased and that this was a direct result of leaving the EU. As only 6.5% of UK school children attend independent schools (Independent Schools Council, 2021), this suggests that the challenge of learning a second language by most school children is compounded by a lack immersion and engagement with the language.

Finally, this leads to a concern about the possible economic, social and personal impact this lack of second language acquisition could have on the UK as a whole and for individuals. Such consequences have been discussed in briefings produced by the House of Lords. Their document states that the decline in second language learning has led to a potential cost to national GDP of 3.5% (Goddard, 2018). A decrease in the uptake of modern foreign language subjects in schools is also showing to influence businesses, with only 34% of businesses rating the foreign language skills of school leavers who are entering the industry as satisfactory (Confederation of British Industry, 2017). The Department for International Trade (2018) has also mentioned the impact of poor language skills in their National Survey of Registered Businesses'. In this 2018 survey, it was found that of the businesses which reported knowledge barriers to exporting, 38% of them listed understanding overseas clients' language or culture as a problem. In businesses with a turnover of over £500k, the percentage that listed this as a barrier increased to 59%. This demonstrates the direct effect on British industry and businesses the lack of UK second language skills has. This is highlighted further by Nick Gibb's (the previous Minister of State for Schools) statement that "In an increasingly globalised economy, it has never been more important for our pupils to be taught modern foreign languages." (Department for Education, 2022).

Although the Government has increased funding to create language hub programmes to help support schools and teachers with extra training (Department for Education, 2022), more need to be done to overcome the poor uptake of second language learning in schools. Possible solutions include increasing the number of hours students spend learning, to make it compulsory from key stages 1-4, and to increase the use of immersive activities. However, with roughly 30% of children in the UK failing to achieve a grade 4 (pass) at GCSE in the core subjects of Maths and English (Burge & Benson, 2023), this seems an unlikely scenario. Moreover, such curriculum changes require time and funds to develop and put into practice.

Therefore, there is room for the implementation of alternative strategies that can make second language learning in schools more effective, without the need for extra curriculum time or funding. This is an on-going historical dilemma, with concerns raised as far back as the 1960s. Asher (1969) highlighted the difficulty American students faced in becoming simultaneously proficient in multiple areas of L2 learning (listening, speaking, reading and writing). He addressed this by proposing a new teaching technique, the Total Physical Response (TPR) based on incorporating physical actions as students learnt.

In the same vein, utilising additional cues during second language learning could aid the acquisition of vocabulary, another challenging aspect faced in UK schools. The use of such cues is one method that could be easily and quickly incorporated into teaching.

Extralinguistic cues for L2 learning

Second language vocabulary learning, without frequent exposure and full immersion to the L2, can be challenging. Cues can be a useful tool employed during L2 teaching to aid in the retrieval process of new L2 words. These cues can be considered *linguistic* (e.g., in the form of L1 translations) or *extralinguistic* (non-language-based). The latter includes picture and gesture cues (Bara & Tirassa, 1999). Although, as emphasised by McNeill (1985) gestures and language are very closely interlinked and share a computational stage, there are many differences in the way they convey meaning. Linguistic communication is linear, segmented and unidimensional in character, whereas extralinguistic communication has global, synthetic and non-hierarchical properties (McNeill, 1992).

During L2 vocabulary learning, extralinguistic cues are believed to be beneficial over L1 translations. Even in the recall of L1 words, accompanying picture cues have been found to be advantageous over only written text during free recall of objects (Paivio & Csapo, 1973). Several theories for this picture effect were discussed by the authors. These explanations were based on the potential differences in the encoding processes involved for the two cue conditions. The key explanation given, that is well supported by the literature, is the Dual Coding Theory (DCT). This theory of cognitive processing proposes that all information is categorised into verbal and nonverbal stimuli and encoded into two separate stores as logogens and imogens respectively (Clark & Paivio, 1991). The pictures with accompanying text condition would have information encoded into both the verbal and nonverbal stores thus increasing the number of encoding routes and aiding recall (Paivio & Csapo, 1969). It is important to note that the DCT refers to verbal and nonverbal stores, into which linguistic stimuli would be encoded into the former, and extralinguistic stimuli into the latter. To avoid confusion, when referring to these stores, this thesis will use the theory's original terms of verbal and nonverbal. However, the term extralinguistic will be used rather than nonverbal to describe cues such as pictures and gestures more appropriately (see McNeill (1985) for the rationale on avoiding the term nonverbal to describe gestures).

The DCT model has also been applied to second language learning with the Bilingual DCT (Paivio & Desrochers, 1980). Compared to the DCT, this bilingual model has separate word logogen systems for different languages. These stores have connections between them, as well as to a common image system. This image system has been of interest to research as it proposes that if two languages have been learnt in separate contexts, then the referential imagens (within the image system) for L1 and L2 may have differences. To investigate this further, Jared et al. (2013) tested picture naming in Mandarin-English bilinguals using culturally biased and unbiased images. The results were supportive of the Bilingual DCT, with culturally biased images in the congruent language being named quicker than in the incongruent language.

Another advantage that has been discussed in using extralinguistic cues in L2 vocabulary acquisition is the development of conceptual links between new L2 words and meanings. As previously discussed, according to the RHM, in early L2 learning there are only weak links between the L2 lexicon and conceptual meanings, with strong L2-L1 lexical links mediating L2 translations. However, research has suggested that mediation through direct conceptual links from the L2 may be available even in the early stages of L2 learning (Sunderman & Kroll, 2006), eliminating the need for the superfluous L1 mediation. Research into learning techniques that can facilitate the early establishment of these conceptual connections is therefore beneficial for L2 learning. By using extralinguistic cues (such as pictures) during learning, these conceptual links have the potential to be strengthened early in the process due to the direct input of sematic information. This could increase the likelihood of semantic mediation of L2 words in translation.

Several studies have investigated if this picture advantage prevails for acquisition of L2 words. Comesaña et al. (2009) compared the effectiveness of L2-picture pairs with L2-L1 translation pairs when learning new L2 words. They found a semantic interference effect in children who were early-stage learners of an L2, but only when taught using pictures. Even though the translation recognition results were similar for L2 words learnt through the different teaching methods, a semantic interference effect was only found with the picture group. This demonstrates that in beginners learning new L2 vocabulary, this effect is dependent on the way the L2 words had been taught, and how picture learning can directly impact on the strength and the development of the conceptual links between L2 and concepts. This effect is also evident in adult L2 learners (Palmer & Havelka, 2010). The use of pictorial cues in learning resulted in greater success in L2 word learning compared to traditional methods of L1 to L2 translations. The application of the RHM model to this research offers a further explanation for this picture advantage over words in L2 memory.

Picture cues compared to gesture cues

Another extralinguistic cue that has frequently been compared to picture cues are gestures. Repetto et al. (2017) tested memory performance using cued and free recall for abstract L2 words that were taught using one of three enrichment conditions: reading, reading and picture cues, or reading and the enactment of gestures. Better recall was found for words learnt with gestures and participants made less errors compared to words learnt with pictures.

This gesture advantage over pictures in L2 vocabulary learning has also been found in young children (Tellier, 2008). This study, which tested French pre-school children, found far greater recall (active knowledge assessment) for L2 words that had been taught alongside gestures than pictures. This effect was persistent in long term memorisation, with the gesture advantage still present after a week without hearing the words. No difference was found in the passive knowledge assessment in which children had to pick the appropriate cue for the L2 word. However, this task was more challenging for the gesture condition as they had to produce the gesture on their own whereas the picture group had a set of images to choose from. Tellier (2008) explains the active knowledge task results through the DCT and suggests that gestures have the advantage of adding a third modality, motor, to the memory trace.

Porter (2012) also tested the effectiveness of gestures in enhancing memory in children for L2 stories. Two stories were read, one with pictures and the other with pictures and gestures. More words were recalled from the story told with gestures than the picture-only story. However, a delayed post-test 2 weeks after learning the stories found that there was greater attrition of the words from the gesture story than the words in the picture-only story. The percentage reduction in words recalled in the gesture story was over double of that in the picture story. Porter (2012) concludes that to maintain the enhanced learning effect that gestures can provide, second language vocabulary should be frequently repeated, revisited and refreshed.

Further research has questioned the presence of the gestural advantage. Andrä et al. (2020) taught children concrete and abstract L2 vocabulary using gestures, pictures or L1 translations. Gesture and picture cues were both advantageous over the no-enrichment condition, but no significant differences were found between the two enrichment conditions in subsequent recall or translation tests. These benefits also persisted up to 6 months after learning, demonstrating the robust long-term retention for words taught using these extralinguistic cues. The equal learning recorded in the gesture and picture learning conditions may suggest that the gesture advantage depends on the word class being taught, as concrete and abstract nouns were used in this study. This, along with Porter's (2012) research, suggests that the gesture advantage in children may depend on several teaching factors (e.g., the frequency of exposure and assessment).

In addition to the empirical research that has frequently found an advantage of gestures over pictures, there are other, more practical reasons, as to why gestures should be incorporated into L2 learning. Firstly, they do not require any pre-planning as pictures do and are far more readily available and practical when teaching. They can also be integrated more naturally within a conversation, by both the teacher and the learner, than pictures. Kelly et al. (2008) highlight this, stating "research in education suggests that teachers can use gesture to become even more effective in several fundamental aspects of their profession, including communication, assessment of student knowledge, and the ability to instil a profound understanding of abstract concepts in traditionally difficult domains such as language." (p.569).

Gestures

Research has explored this gesture advantage further to understand why gestures produce superior learning effects compared to other extralinguistic cues, and how best to enhance the effect. To produce a coherent analysis of these different explanations and explore these advantages in depth gestures must be looked at more closely.

The different functions, uses and movements involved in gestures can be used to categorise them into specific types. The names of these categories have varied throughout the years, but McNeill (1992) identified four main types of gestures. Iconic gestures are those that provide related semantic information to the accompanying speech (see Kandana Arachchige et al. (2021) for an overview of how these gestures have been defined across the literature). They are used to aid communication by providing additional visual cues and information complimentary to the speech. Similarly, metaphoric gestures also use imagery, but they are used to represent abstract concepts. Deictic gestures are any form of pointing, showing or demonstrating usually using an outstretched finger but can also include the use of other body parts including the head or chin (McNeill, 1992). These gestures typically make use of the gesture space, rather than pointing at a specific entity. Lastly, beat gestures do not have a distinct meaning on their own but display movement patterns that fit with the rhythm of speech. These gestures are typically performed with fingers or hands and, unlike deictic gestures, do not use special gesture space but are executed wherever the hands find themselves.

Gesture uses: Memory, comprehension and speech integration

As with pictures, gesture cues have been found to be useful tools across a range of learning settings, including L1 verbal recall (Frick-Horbury, 2002), novel word learning (Macedonia & Von Kriegstein, 2012), comprehension (Dargue et al., 2019) and improving perceptions and social evaluations of non-native speech (Billot-Vasquez et al., 2020). The role of iconic gestures specifically in the comprehension of speech has been investigated at length (for an overview, see Kandana Arachchige et al., 2021), with these being found to play a significant role (for review see, Hostetter, 2011).

One explanation for this benefit of gestures in comprehension is the strong interaction between speech and gestures in the brain (Willems & Hagoort, 2007). This can be explained further through the integrated-systems hypothesis proposed by Kelly, Özyürek and Maris (2010). In their research, action primes were used followed by gesture-speech pairings that either contained no related information to the prime, a congruent baseline with gesture and speech related to the prime, or partially related information to the prime (in the speech or gesture). The latter condition also varied in the degrees of congruency. The target videos were either a weakly incongruent condition (e.g., 'cut' instead of 'chop') or a strongly incongruent condition (e.g., 'twist' instead of 'chop'). These differing levels of incongruency were manipulated across both speech and gestures. Greater accuracy and faster reaction times were observed in the congruent baseline condition. Additionally, weakly incongruent pairs produced fewer errors than strongly incongruent pairs across both gestures and speech, suggesting that the influence of modality on comprehension was comparable across the two. In a further experiment, Kelly, Özyürek and Maris (2010) altered the task so that only the verbal information was the target of the videos. Even when participants were not told to respond to the gestural information and only needed to identify if the speech matched the prime, there were still greater error rates with strongly incongruent gestures compared to weakly incongruent ones. The integrated-systems hypothesis therefore predicts gesture and speech interactions to be both mutual and obligatory.

The integrated-systems hypothesis therefore suggests that this integration has automatic processing properties, with gestural information influencing the unrelated task unavoidably. Further research from Kelly, Creigh and Bartolotti (2010) has investigated this using both behavioural and electrophysiological measures. In order to empirically test automatic processing, participants were presented with semantically congruent and incongruent gesture-speech pairings. These pairings also varied in their gender congruency, with the actor performing the gestures and the speakers voice either being congruent or incongruent. The gesture-speech pairings were presented in a Stroop-like task and the interference effects were measured. Participants' task was to indicate if the speech was male or female. In this way, the gesture-speech congruency was not relevant to the participants' task. Reaction times were measured to determine if, despite having no relevance to the task, gesture-speech incongruency would affect the speed of responses, indicating automatic gesture-speech integration. Indeed, longer reaction times were recorded for semantically incongruent gestures than for congruent, demonstrating the semantic interference effects. For the electrophysiological measure, event related potentials (ERPs) were recorded during the task. Previous literature has identified the N400 component as being associated with semantic processing (Hinojosa et al., 2001), with a smaller N400 component demonstrating greater semantic integration and processing (Kutas & Hillyard, 1980). Kelly, Creigh and Bartolotti (2010) reported a larger N400 for incongruent gesture-speech pairs than congruent, demonstrating the automatic nature of the integration of the two modalities. The behavioural findings produced from the gesture-speech Stroop-like task in this research have been replicated (Zhao et al., 2018, 2021). These studies also provided evidence that the left inferior frontal and posterior temporal gyrus causally contribute to the semantic integration of gesture and speech.

However, Kelly, Creigh and Bartolotti (2010) discussed additional findings that suggest that this integration may not be an exclusively automatic process. The integration was also affected by gender congruency across the speech and gestures. For the reaction time data, stronger gesture congruency effects (larger differences in reaction times) were found when the gender was the same across the speech and gesturer compared to different. This suggests a level of sensitivity to the context of the task on the automatic integration.

Additionally, this level of control over the integration has been demonstrated in another study that varied the explicit instructions given to participants about whether the gesture and speech information belonged together or not (Kelly et al., 2007). Similar findings were also found in a study by Holle and Gunter (2007), with the input of non-meaningful hand movements disrupting the integration of meaningful gestures and speech in sentence processing. This suggests that these situational factors can influence the obligatory process of gesture-speech integration.

Speech-gesture vs speech-action integration

Gesture-speech integration is thought to have similar properties to action-speech integration (Willems & Hagoort, 2007). But the question remains as to whether gestures are unique in their connections to speech during language comprehension, or if this special relationship also applies to actions. Communicative gestures and manual actions that accompany speech are obviously very similar- they often share the same purpose and display a similar pattern of movement, with some researchers describing gestures as simulations of actions (Hostetter & Alibali, 2008).

Research has looked to determine if the integration for speech-gesture pairings is different to that in speech-action pairings in terms of processing speed and the communicative benefits. Kelly and colleagues (2015) extended this research to explore if speech-gesture pairings are integrated in the same way as speech-action pairings, or if gestures are unique in some way in their ability to be integrated with speech. Multimodal stimuli were presented with action or gesture video clips paired with either congruent or incongruent speech. A written word prime preceded the video clip and was either related or unrelated to the video auditory and/or visual information. Participants' task was to determine if the visual or auditory information were related to the prime, with half of participants focussing on the visual information and the other half on the auditory information. The results found that for both visual stimuli modalities (gestures and actions), in both the visual and auditory target conditions, congruent pairings had quicker and more accurate responses than incongruent pairings. This difference, however, was greater for the speech-gesture stimuli than for the speech-action stimuli. This suggests that gesture and speech are more closely interlinked during speech processing than actions and speech. Additionally, within the visual target task, responses to speech-action stimuli were quicker and more accurate than to speech-gesture stimuli, suggesting that actions are processed more easily than gestures.

The authors link these findings to Paivio and Csapo's (1973) DCT, with actions appearing to have a 'richer visual code' – therefore are processed more easily. Actions are said to be more visually complete and informative (less ambiguous) than gestures. The findings of gestures being more involved in speech processing were also explored. Kelly et al. (2015) suggested that although gesture appeared less visually informative than actions (having slower reaction times and lower accuracy in the visual target task), gestures are proposed to be more communicatively informative (displaying greater semantic congruency effects). The authors explain that "viewers may generally assume that gesture, more than action, is information *meant to* accompany speech, and this may increase their attention to it." (Kelly et al., 2015, p.522). This idea that the intention behind the use of gestures can play a part in how they are integrated with speech, is supported by the research mentioned previously on the context given for the task (Holle & Gunter, 2007; Kelly et al., 2007).

Embodied representations

Another explanation for the strong impact of co-speech gestures on cognitive processing is through embodied cognition. This perspective comes from the idea that our bodies can have a strong influence on our mind and vice versa, with many physical actions (including gestures) engaging the perceptual and motor system, which can aid cognition (Macedonia & Repetto, 2017; Madan & Singhal, 2012). The engagement of the motor system during memory tasks using gestures becomes particularly evident when additional unrelated motor tasks are also completed. In Halvorson et al.'s (2019) study, completing an unrelated motor task did not affect the gesture advantage in the memory of learnt phrases if the movement of the motor task was the same during encoding *and* retrieval. Inconsistences in the motor task during encoding and retrieval caused the gesture enhancement on memory to diminish.

Explanations of the gesture advantage for L2 vocabulary learning

The body of research previously discussed demonstrates the integration of gestures with accompanying speech and the communicative benefits iconic gestures can have. The application of this approach and the development of embodied lexical representations in L2 learning has also been discussed (for an overview see, Macedonia, 2014). Given this, the benefits of using iconic gestures during L2 vocabulary learning have been explored.

In the previous section 'Picture cues compared to gesture cues' (p.30), I briefly discussed the gestural advantage for L2 word learning when compared to pictures, before moving on to explain the benefits and uses of gestures in broader contexts. However, much empirical research has explored the extent of this advantage under different conditions and when compared to other cues. The first empirical study to do so was conducted by Allen (1995), who found greater recall and long-term retention of short L2 sentences when accompanied by gestures, compared to when there were no additional cues. Since then, many other authors have added to the literature (for reviews see, Gullberg, 2008; Macedonia & Von Kriegstein, 2012). In the following section, this further research will be explored, along with the theories and explanations offered to aid the understanding of the gestural learning effect found. The gesture-speech integration highlighted in the previous literature in the earlier section (Gesture uses: Memory, comprehension and speech integration, p.32) may contribute to the gestural benefit displayed in L2 learning. However, there are other possible explanations proposed in the literature that could also explain this gestural advantage including the DCT, the Motor trace theory and self-involvement theory.

One theory that has been applied to understand the advantage of gestures in L2 vocabulary learning is DCT. As mentioned previously, this theory can explain the general advantage of the use of extralinguistic cues, but it can also be extended to explain why gestures may have greater benefits to pictures. The DCT predicts that the more modalities available for encoding, the more memorable the learning will be (Clark & Paivio, 1991). Picture cues would provide auditory and visual coding routes, but gestures would provide these two routes in addition to a third motor route (Huang et al., 2019).

Huang et al. (2019) suggested that, when applied to the use of gestures in L2 learning, the theory would assume that gestures do not necessarily have to be meaningful. The DCT would suggest that other gestures could also be helpful, regardless of their semantic content, as multiple sensory modalities will create a richer memory trace. They investigated this by testing the effectiveness of low and high idiosyncratic gestures in L2 word learning. Low idiosyncratic gestures were defined as those that are classically iconic and directly link with word meaning, whereas highly idiosyncratic gestures are not obviously iconic with any word and would need to be paired with word meanings to provide information. The use of these gestures was compared to a gesture-free condition, and recognition was measured using a 4-AFC task. The two gesture conditions did not differ in recall, however they both differed significantly from the no-gesture condition. Huang et al. (2019) conclude that these results support the DCT and suggests that, in a simple learning environment where items are unlikely to be confused, even gestures without initial iconic meaning can be helpful for learning.

Other research has produced similar findings, with So et al. (2012) reporting that beat gestures, which lack semantic information, can also be beneficial to memory- being just as helpful in aiding recall in adult learners as iconic gestures. However, they did not find the same results across adults and children, with children only displaying the benefits of iconic gestures, and beat gestures producing similar results to the no-gesture condition.

Huang et al. (2019) suggested that an alternative to the DCT as an explanation for both sets of research above is that gestures that do not contain initial iconic meaning can have meaning applied to them during learning, as long as these gestures do not have any pre-held meanings. This could then explain why So et al. (2012) found differences in gesture type in children, as they may not be mentally advanced enough to perform this strategy mid-learning.

The premise of the DCT has also influenced other theories used to explain the gesture advantage, including the motor trace theory. This theory proposes that the motor encoding created by gestures, and particularly the enactment of gestures, leaves an enhanced motor trace in memory (Engelkamp, 2001; Engelkamp et al., 1994; Engelkamp & Zimmer, 1985). This motor trace then becomes part of the word's representation, enhancing retrieval with the multimodal representations.

Performing the gesture is a key component of the motor trace theory to help to establish and strengthen these representations and this was described by Engelkamp as the 'enactment effect' (Repetto et al., 2021). The role of this enactment effect has been largely discussed and tested, when it comes to the use of gestures in learning. The production of gestures has found to be advantageous in; L1 recall in young children (Tellier, 2007); in L2 learning over viewing pictures or outlining pictures with finger (Mayer et al., 2015); in long-term memory of L2 words compared to audio-visual cues (Macedonia & Klimesch, 2014). The formation of these motor traces was investigated in Macedonia and Mueller's (2016) study which found that L2 words learnt through self-performed gestures prompted activity in key motor cortical areas during recognition. The development of extensive sensorimotor networks (motor traces) during enactment of gestures was concluded to aid memory.

Indeed, this enactment effect makes sense when considering the strong embodied view of recognition memory (Repetto et al., 2021). Research has found that enacting words through full body movements in physical exercise leads to greater learning compared to part-body movements in enacting gestures (Mavilidi et al., 2015). However, many empirical papers have still found gesture advantageous despite participants not performing the gestures (Huang et al., 2019; Kelly et al., 2009; So et al., 2012). Huang et al. (2019) take this to suggest that the motor trace can still have an effect even when just viewing gestures. This implies that the enactment of gestures is not essential but can further enhances the effect. It may also be that imagining oneself performing the gesture can stimulate the motor traces. Positron Emission Tomography (PET) studies have found that the same motor cortices were activated during performing an action and imagining performing the action, but to a lesser extent for the latter (Nilsson et al., 2000; Nyberg et al., 2001). Considering the similarities between actions and gestures, imagining performing a gesture may also stimulate motor traces to a similar extent as performing.

Another perspective that places importance on the performance of gestures is the self-involvement explanation. This theory predicts that participants become more involved in the learning process by reproducing the gestures and that this increases their attention to the learning material (Helstrup, 1987). This explanation if applied alone, would suggest that any movement that engages the learner in the learning process can be just as effective as meaningful, iconic gestures. However, a great deal of research has found this not to be the case, for example Mayer et al. (2015) found that performing gestures still produced greater learning than using hand movements that traced images. Similarly, the self-involvement perspective can’t explain differences between two gesture conditions that both involve performance of the gesture, including comparisons of meaningful and meaningless gestures.

The importance of semantic relatedness in gesture-supported L2 learning

Several studies have investigated the importance for semantically meaningful gestures in the gestural advantage in L2 word learning. Research in this area demonstrates that the gestural advantage goes beyond simply creating a more memorable learning environment through hand movements, and that the semantic information is an important factor. Research has compared the use of meaningful and unrelated gestures in L2 vocabulary learning. Kelly et al. (2009) tested different teaching methods with a population of English native speakers whilst learning Japanese L2 words. The Japanese verbs were learnt audio-visually and by performing either congruent or incongruent gestures. It was found that words learnt through congruent gestures had greater recall than those learnt using incongruent gestures or just the audio-visual stimuli alone. However, the incongruent gestures being used were still iconic and meaningful, but for other words being learnt. This may therefore have caused interference effects, as, according to the motor-trace theory, these gestures will have already had strong motor trace connection to other words. This research is therefore not comparable to that conducted by Huang et al. (2019) who used unfamiliar gestures.

Further research supports the importance of semantic gestures for L2 learning. Macedonia et al. (2011) also found superior learning effects of congruent iconic gestures over meaningless gestures, and these were consistent over time. García-Gámez and Macizo (2019) compared the use of congruent, incongruent and meaningless gestures, as well as having a no-gesture condition. Greater learning was achieved for L2 nouns and verbs when learning was conducted through congruent gestures. Interestingly, they found that incongruent and meaningless gestures negatively impacted learning, with learning in these conditions below that of the no-gesture condition. This re-emphasises the automatic integration of speech and gestures and the detrimental effects that conflicting gestures can have. Additionally, it suggests that meaning cannot be applied during learning for all meaningless gesture, as research has suggested (Huang et al., 2019; So et al., 2012).

Disambiguation account

The research mentioned previously, showing the need for congruent gestures, could be taken to suggest that semantically congruent gestures can provide extra meaningful, unambiguous contextual information that benefits memory. The use of gestures has been found to assist in verbal ambiguity in L1 communication (Holler & Beattie, 2003). In the L2 literature, the majority of studies present isolated action words, which are semantically underspecified (e.g., push – this could refer to different contexts, push a button, push someone over.). As such, gestures may boost learning by naturally disambiguating the context of the word.

The disambiguating properties of gestures have been explored, with Holle and Gunter (2007) investigating the use of gestures in disambiguating speech, using EEG to monitor the N400 component. Participants were presented with sentences that contained an initial unbalanced homonym in the first half, followed by disambiguating speech in the second half of the sentence. Iconic gestures were presented by the speaker alongside the homonym and were either supportive of the dominant or subordinate meaning. The N400 was smaller (indicating greater integration) after a congruent gesture had been presented than when an incongruent gesture was presented. This demonstrates the use of gestures as disambiguating cues for speech, as the gestures strongly influenced the word meanings that become activated.

This therefore suggests that meanings are an important aspect of the gesture advantage as they can be used to disambiguate a learning situation by creating a clearer and thus more memorable environment.

Privileged access account

It is likely that several possible mechanisms may account for the gestural advantage. Many of the afore mentioned theories hold similar ideas-that iconic gestures are special and can provide an advantage over other cues, whether that be from the development of specific motor-traces or via embodied representations. I will combine the key components of these perspectives and throughout this thesis, and collectively term them as 'the privileged access account'. This assumes that gestures may provide privileged access to action representations and motor traces during learning, that are not provided by other learning methods. These representations can deepen the sensory motor image that accompanies the L2 word, capturing the meaning in an embodied way making them more memorable. Although this exact term is not frequently used, the privileged access of gestures has been mentioned in the previous literature when compared to other cues, "in our cognitive system, gesture holds a specific privileged status compared to sensory and simple motor modalities" (Repetto et al., 2017, p.9).

Multiple cues

Another factor that could influence the benefits of including gestures during learning is the use of multiple cues. During L2 acquisition, the more cues available to the L2 word, the greater the learning (e.g., Sueyoshi & Hardison, 2005). This is regardless of what form the cues take but implies that multiple cues will create more traces for encoding the L2 word, thus providing a more memorable environment than fewer cues. This proposed advantage is relevant to the L2 vocabulary learning research on gestures, as a large proportion of the past literature included gestures as an additional cue. For example, Porter (2012) found a gestural advantage in the recall for L2 words in stories, but gestures were presented alongside pictures. The gesture advantage found could therefore be explained through the additional cue being provided during learning. To accurately test the benefits of gestures in L2 word learning, the number of cues needs to be controlled for across learning conditions.

Similarly, Dual Coding Theory (mentioned previously in 'Explanations of the gesture advantage for L2 vocabulary learning', p.35) could also explain the gestural benefits displayed in some of the past literature. The theory suggests that the involvement of different cue modalities for routes to encoding into both the verbal and nonverbal systems may also contribute to advantages (Clark & Paivio, 1991). In previous literature (e.g., Kelly et al., 2009), gestures have been used as an additional cue alongside L1 translation (extralinguistic and verbal cues) compared to L1 translations presented as the only cue to the L2 word (verbal only cue). Thus, the benefits of the gesture learning condition over L1 translations could be explained through the DCT as this condition provided encoding into both the verbal and nonverbal stores.

Hence, the Disambiguation and Privileged access accounts will form the theoretical background for my thesis, which aims to explore the mechanisms underpinning the gesture advantage in L2 word learning. The use of multiple cues and encoding routes into different stores will also be a factor that will be considered in the following series of studies.

Summary of open questions and research aims

There are a number of key open questions that still remain within this research area. I will summarise these before describing how they will be addressed through my series of experiments.

Presentation of gestures alongside other cues

Although the literature into this gesture advantage during L2 learning is extensive, there are several common methodological issues. Firstly, to the best of my knowledge, all empirical papers bar one, use gestures during learning that are accompanied by L1 translations. Other than Tellier (2008), no research appears to have taken advantage of the fact that iconic gestures already display semantic information, and thus do not need to be accompanied by the L1 translation when used as cues in L2 vocabulary learning. However, there has been naturalistic work in which gestures have been effectively implemented into L2 teaching without the need for the inclusion of the L1 (see McCafferty & Stam (2009) for an overview). For example, Smotrova and Lantolf (2013) observed the use of gestures to aid explanations of written L2 lexical concepts within a classroom setting during interactive discussions. Additionally, it must be noted that while empirical research has previously paired gestures alongside only L2 speech, these studies were not investigating L2 word learning specifically. Rather, the use of gestures was explored in assisting L2 narrative recall (Lin, 2021) and listening comprehension (Sueyoshi & Hardison, 2005) in intermediate and advanced learners.

Within the previous gesture L2 learning literature, the learning sessions typically involved three cues, L1, L2 and gesture. Based on the RHM, a combination of just L2 and gesture could potentially trigger a faster learning process than presentation of all three cues combined. This is because inclusion of the L1 would be expected to facilitate lexical connections between L2 and L1 words, whereas this laborious L1 detour could be bypassed when only a gesture accompanies the novel L2 word, as the more direct conceptual route would be facilitated.

Can gestures fast-track L2 word learning when used in isolation?

Limited outcome measures

Another key question that remains stems from the past literature's narrow variation in the outcome measures being used. Research in this area has a heavy emphasis on explicit tasks, with research often implementing either cued and/or free recall as the behavioural measure. As a result, it is not clear from these studies if the advantageous effect of gestures can also extend to tasks measuring automatic processing, or whether the advantage is limited to controlled, conscious retrieval of the L2 words.

As mentioned in the previous section 'Automaticity in language processing', p.2, automatic processing is a key characteristic of language use and without its development in L2, learners will be limited in their L2 proficiency. In order to reach high L2 proficiency, automatic processing abilities must be developed. It is therefore important to test if the gestural advantage can extend to tasks measuring this type of processing, under conditions that are reflective of real-world conditions.

The limited behavioural outcome measures implemented in the L2 gesture learning research becomes particularly apparent when compared to the range of techniques and tasks used to measure automatic processing within the novel word learning literature (see 'Semantic integration', p.8). By incorporating a range of behavioural measures, including masked and unmasked priming, eye tracking and Stroop-like tasks, the automatic processing of newly learnt novel words has been assessed. However, using such automatic processing tasks as a behavioural measure have not been implemented in the L2 gesture learning research. If the gesture advantage could still be found in a speeded, implicit task, it would suggest that gestures can facilitate the automatic processing of newly learnt L2 words as the effect would be demonstrated in rapid, unconscious conditions.

Additionally, a task that is implemented under such automatic conditions would be able to measure the development of direct conceptual links between the L2 words and concepts. Although a cued/free recall task can demonstrate if there are overall benefits to learning through a gesture learning method, it cannot specifically identify the use of direct semantic connections due to the timeless nature of the task.

Does the gestural advantage persist under implicit, automatic processing conditions?

Uncertainty over which mechanisms cause the gesture advantage

Finally, it is not clear from the research which mechanisms are the key drivers of the gestural advantage. As discussed previously, a number of possible explanations have been mentioned within the literature but few attempts have been made to directly test the differing predictions of these explanations against each other in a systematic way.

What mechanisms are involved in the gestural advantage?

How each open question will be addressed in my research

Firstly, to test the prediction from the RHM that gestures presented in isolation can facilitate the more direct conceptual link, one aim of this research is to assess if the gesture advantage can persist when used as the sole cue to L2 words. This will be the first study to test the use of gestures as the only cue for L2 word learning in adults. Additionally, to truly test the application of the RHM and assess if the use of gestures alone can fast-track learning, the development of conceptual links needs to be measured. To do so a new outcome measure that can test for semantic effects of the newly learnt L2 words needs to be used.

The implementation of such a task will also address the shortcomings of the previous literature in their use of limited, explicit tasks. The following series of experiments therefore used a new outcome measure that aimed to test the semantic integration of L2 words in an implicit, speeded environment. This will allow the gestural advantage to be tested under more automatic processing conditions.

As well as the afore mentioned aims to test L1 in isolation and use a new outcome measure, this research also intends to address the mechanisms that drive the gestural advantage. I aimed to explore the possible explanations mentioned previously, namely the privileged access and disambiguation account. In order to do so, the learning cues used throughout the following studies were systematically manipulated to understand the involvement of these mechanisms, as well as the possible influence of multiple cues and multiple encoding routes (DCT), in the gesture advantage.

STUDY 1 – TESTING THE EXTENT OF THE GESTURAL ADVANTAGE

Introduction

The aim of Study 1 is to test whether gesture-aided learning leads to deeper semantic retention of newly learnt L2 words than traditional translation learning methods. This learning was measured using both an explicit cued recall task as well as a speeded, implicit task. Additionally, this research aimed to test this gesture benefit when no other cues to the L2 words are provided.

As discussed previously in the general introduction (Extralinguistic cues for L2 learning, p.28), much research has been conducted on the use of gestures as cues to L2 word learning. Iconic gestures have been found to be beneficial in aiding this learning, outperforming other cues such as L1 translations (Huang et al., 2019), pictures (Repetto et al., 2017), meaningless gestures (García-Gámez & Macizo, 2019) and incongruent gestures (Kelly et al., 2009). This gestural advantage has also been evident in children (Tellier, 2008).

However, from the past literature, it is not clear how this gestural advantage comes about. It may be that the advantage is just a result of a greater episodic memory trace provided through gestures. But conclusions cannot be made as the previous research used limited outcome measures that were heavily focused on explicit tasks. Therefore, it cannot be determined if the gesture learning effects found only represent greater explicit, episodic memory or if they can also extend to reflect automatic processing. As discussed previously in the 'Automaticity' section of the general introduction (p.1), automatic processes are fast-acting and do not require conscious control or effort (Schneider & Shiffrin, 1977). Therefore, in order to test such automatic processing, a task would need to test learning under these automatic conditions.

For the development of Study 1, I also considered how gestures have been presented in the previous literature. When applying Kroll and Stewart's (1994) RHM to this research, it would be assumed that, due to the strong lexical links created between L1 and L2 early on in the learning process, any learning condition that includes the L1 translations would facilitate the strengthening of these links. In the early stages of language learning, concepts for newly learnt L2 words are accessed via the lexical links to the L1, and then the strong, developed conceptual links from the L1 to the meanings are used. However, if a learning condition did not provide access to the L1 and instead only contained semantic information on the word, the lexical links between L2 and L1 would not be facilitated, restricting their development. Instead, the model would suggest that, given they become the only link between the L2 and meaning, the direct conceptual links from the L2 would be taken and thus become more developed.

However, the use and implementation of gestures within learning has been limited in the past literature. To the best of my knowledge, all empirical research bar Tellier (2008), have used gestures alongside the L1 translations as cues in L2 vocabulary learning (but see McCafferty & Stam (2009) for naturalistic work in this area). Consequently, this research may not be a true reflection of the full extent of the gesture advantage. The use of gestures alone, and the possible benefits of doing so, are yet to be explored.

Another model mentioned previously that is important to consider within this research is the CLS (McClelland et al., 1995). The CLS states that a period of deep consolidation is key for the lexicalisation of newly learnt words. Research has found varying results on whether such a consolidation period is necessary or not for newly learnt words to display automatic processing on a lexical and semantic level. However, on the whole, a consolidation period does appear to be beneficial to learning. Indeed, after several nights rest the CLS would suggest that a shift to more automatic processing can occur through the neocortical system.

A large proportion of the past literature on the use of gestures during L2 word learning did indeed implement a consolidation period (e.g., Kelly et al. (2009) tested 2 days and 1 week after learning and Macedonia & Klimesch (2014) assessed learning over a period of 14 months). Typically, this research has also tested immediate learning effects on the same day as learning. Yet some previous research only implemented testing straight after learning with no other delayed testing (e.g., García-Gámez & Macizo, 2019; Huang et al., 2019). This does not allow for consolidation which, according to the CLS, would benefit the integration of newly learnt words. The benefits of a consolidation period for learning have also been evident specifically with the use of gestures. Cook et al. (2013) found that when children learnt mathematical problem-solving skills through a speech plus gesture learning condition, they showed improvements in scores after a 24-hour delay for memory consolidation, compared to testing immediately after learning. In comparison, no benefits to test scores after the delay was found for the speech only learning condition. This highlights the importance of a consolidation period in L2 learning, as assessment after this time could allow for the display of an even greater gestural advantage.

Given these considerations and gaps identified in the current literature, Study 1 had several clear goals. Firstly, in order to test automatic processing of newly learnt words, reflective of deep semantic learning, a new outcome measure was implemented in the assessment session. This new task was adapted from the gesture speech-integration task from Kelly, Creigh and Bartolotti's (2010) study to use as a measure for the semantic integration of L2 words. The task, which is a variant of the classic Stroop procedure designed to test automatic processing (Stroop, 1935), was used to demonstrate how gesture and speech are automatically integrated. Within the task, video stimuli are presented of an actor or actress performing an iconic gesture. Each video clip is displayed alongside speech that names an action word, and this speech is either semantically congruent or incongruent with the gesture presented. Additionally, the gender of speech and actor in the video clip is either congruent or incongruent. The participants' task is to indicate whether the gender of the speech is male or female. Despite the semantics of the spoken language not being relevant to the participants' task, semantic congruency effects were found suggesting automatic integration of the gesture and language information. Such a Stroop-based task could serve as an implicit outcome measure for assessing L2 semantic knowledge in speeded conditions. If semantic knowledge of the L2 words has been integrated, then semantic congruency effects should be found when using L2 speech in the task. As gestures are being used during learning, the task was also adapted to use action video clips instead of gestures. This new assessment measure was termed the 'action-speech integration task'.

Before implementing the task as an outcome measure for L2 learning, it first needed to be tested to ensure that the same integration effects could be found with the action stimuli. As discussed in the previous section (Speech-gesture vs speech-action integration, p.34), Kelly et al. (2015) used a similar task measuring gesture/action speech integration with written primes. Although both speech-gestures and speech-action pairs produced semantic congruency effects, it was greater for gestures than action. This highlights the necessity of checking that when actions are used in the integration task, strong semantic congruency effects are still present. Pilot studies testing this new assessment measure were conducted before the use in Study 1.

The second aim for Study 1 looked to apply theory from the RHM in exploring the potential of gestures to be used as the sole cue in L2 word learning. In all previous empirical research bar one, gestures have been accompanied by the L1 translation, however given the semantic information provided by gestures, these may not be necessary. Therefore, during learning gestures were presented as the only cue to the L2 words, with only limited exposure to the L1 before learning. The use of gesture cues was compared to L1 translations as they are the most common cue provide during learning in a classroom setting.

Finally, taking into consideration the varying findings, discussed previously, on when this automatic like processing and integration into the mental lexicon starts to develop in newly learnt words, Study 1 included two assessment sessions. The first was conducted one day after learning took place. This was for practical reasons to ensure that the first session, which also included the learning phase, did not become too long. Secondly, this was based on the research on CLS that, for the most part, indicates that a deep consolidation period of sleep can aid lexicalisation. This would therefore increase the likelihood of finding good levels of learning and evidence of semantic integration, particularly when considering the benefits of consolidation when using gestural based learning (Cook et al., 2013). The second assessment sessions took place one week after learning. This was to gain insight into the effects of a longer consolidation period, as well as measuring whether the gesture advantage persists over time.

Aims for Study 1

To recap, to assess whether gesture-supported learning results in deep semantic understanding of L2 words, participants completed an implicit reaction time task, in the form of an action-speech integration task (modelled after Kelly, Creigh and Bartolotti, 2010). I also assessed explicit learning in the form of a L2 to L1 translation task. To assess whether gesture leads to only short-lasting or also longer-lasting benefits, learning was assessed both 1 day after learning, as well as 7 days after learning. The three key questions being asked are highlighted below:

i. Is the gesture advantage still evident in a speeded, implicit task?
ii. Are gestures still beneficial in isolation?
iii. Does the gesture advantage persist over time?

Method

Participants

Participants signed up via the University of Hull's SONA Participant Sign Up website and were paid with either 2.5 research course credits and a £10 Amazon voucher, or a £20 Amazon voucher if course credit was not needed. The voucher reward was only awarded for completion of the full study. If completing for course credit, partial credit could be awarded. Data collection took place from April to August 2021.

In total, 59 participants enrolled in the study, however four of these did not complete the first learning stage, one had incomplete learning data, four completed only the learning stage and no assessment, two completed the learning and the first assessment stage but not the final session and one participant completed the assessment sessions several times and was therefore removed. Sessions completed 1 day late (2 days after learning for day 1 assessment and 8 days after learning for day 7 assessment) were accepted, but any later were deemed as too late. Participants were not informed of this leeway when starting the study to help ensure that the participants completed the sessions on the correct days. One participant completed the day 1 assessment 3 days after learning and was therefore excluded from analysis.

The remaining participant group of 46 that completed the study contained 35 females and 11 males. The mean age of participants was 26.78 (SD = 8.87). Of these participants, 19 received course credit and the voucher and 27 received just a voucher. Out of the 46 participants, English was the first language for 32 participants. For the other 14 participants, 12 stated English as a fluent second language. Although the remaining two participants did not report English as a second language (this may have been an error), it is fair to assume that they are proficient in English as they were either a student or staff member at the University. None of the participants had any previous knowledge of any Chinese vocabulary prior to participation in the study. In a self-reported question on personal motivation to learn Chinese, 24 participants (52.17%) stated that they were personally motivated.

Design

The translation task used a 2 x 2 factorial design, using the two within-subject factors Learning Method (gesture, translation) and Day (1, 7). For the action-speech integration task, a 2 (learning method) x 2 (day) x 2 (semantic congruency) x 2 (gender congruency) design was used.

Learning Assessment Methods

Explicit learning assessment: translation task

In order to assess learning, two assessment methods were used. The first was an explicit assessment that involved a simple translation task from the Mandarin words to English. This task was used as a baseline measure of explicit learning that could be comparable to the cued recall tasks used in the previous literature.

In a random order, all Mandarin words that the participant had learnt were presented individually on the screen in pinyin with a text box underneath (see Figure 5). The participant would hear this word said by both the female and then the male speaker. Participants then typed what they thought was the correct English translation into the text box.

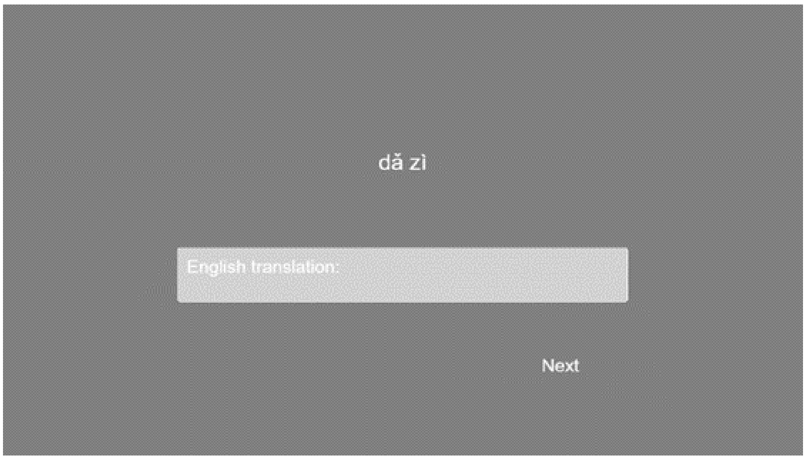


Figure 5. The translation task

Implicit learning assessment: action-speech integration task

The second method was an implicit form of assessing Mandarin word learning. This used an adapted version of the gesture-speech integration task from Kelly, Creigh and Bartolotti (2010). Rather than gesture video clips, our adapted task used actions in the implicit assessment because gestures were already used in the learning phase. Thus, instead of Kelly's gesture-speech integration task, an action-speech integration task was used to measure learning. In this way, the assessment phase videos were novel for all words, for both learning conditions. Additionally, by using different video stimuli in the assessment phase, the generalisation of learning was able to be tested, as similar (but not identical) cues were used in the assessment.

Pretest: validation of paradigm and stimuli set. This action-speech integration task was first piloted, using English speech, both in native English speakers[1] (n = 38) and in non-native, fluent English speakers (n = 40). In both, a significant semantic congruency effect was found, with semantically incongruent trials having longer reaction times than semantically congruent trials (native English speakers – Cohen's d = 0.41, mean difference = 10.77ms; non-native – Cohen's d = 0.13, mean difference = 19.09ms. See Appendix A, p.164 for mean reaction times). This therefore suggests that the action and spoken information are being automatically integrated in both native as well as advanced L2 speakers of English. This task was therefore used in this study, using Mandarin action words, as an implicit assessment of semantic integration.

Within this action-speech integration task, the keys used by participants to indicate a male or female voice were randomised. During the instructions of the task, the keys were presented in the keyboard mapping form to show the spatial arrangement visually to make them easier to remember. To ensure that participants understood the instructions, before starting the task a short practice round was completed using two Mandarin action words that were not part of the learning set.

In each trial of the task, an action clip played and after a 200ms delay, participants heard one of the 24 Mandarin action words. The spoken word either matched or differed to the action being performed. The gender of the voice was either congruent or incongruent to the gender of the actor in the video clip. Each participant had to identify the gender of the voice, by pressing either the F or J cursor key on their keyboard. Participants' reaction times were measured and recorded during the experiment. After responding, a blank screen appeared for 1000ms. Then a fixation cross appeared in the centre of the screen for 500ms before the next video clip began. In the case of an incorrect response, feedback was presented for 500ms followed by the blank screen.

The order of the two assessment tasks were not counterbalanced, with all participants completing the translation task first. This was because the implicit assessment task may have the potential to improve performance in the translation task as in half of the trials, the action videos are paired with semantically congruent speech. This could aid additional learning for word pairing that may not have been correctly identified otherwise. Additionally, to test long-

[1] This was reported in my Masters dissertation.

term retention of newly learnt L2 words and to allow for the greatest effects of consolidation, each assessment session (translation task and action-speech integration task) took place one day after learning as well as 1 week later.

Stimuli

Mandarin was chosen as the L2 as it was less likely that participants, drawn from a pool of UK undergraduate students, would have any prior experience of this language compared to European languages, which are predominately taught in UK schools. This therefore minimised the number of people who would be deemed ineligible to take part.

Mandarin is the most commonly spoken dialect of the Chinese language spoken in China. Pinyin (hànyǔ pīnyīn) is the form of transcribing and spelling out Mandarin Chinese sounds using the Latin alphabet. Mandarin is a tonal language, as pitch is used to distinguish meaning across words. These so-called lexical tones are expressed in the pinyin text using accent marks. There are four tones, and thus four accent marks: ā, á, ǎ, à. The first, ā, produces a tone that starts and remains high pitched. The second, á, uses a tone that starts with a medium pitch and rises. The third, ǎ, emits a pitch that falls before rising to a higher pitch. Finally, the fourth, à, indicates a pitch that drops from high to low.

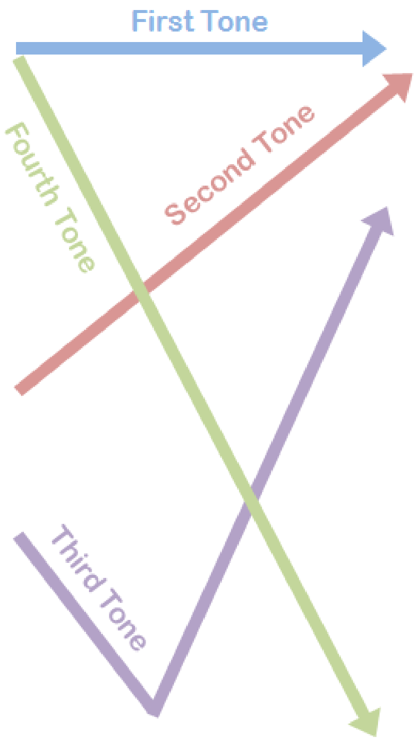


Figure 6. Visual display of the pitch changes for each of the Mandarin tones (Mandarin Tutor, 2024)

Given the complexity of Mandarin with these different tones and the considerable contrast to other non-tonal languages, when selecting the Mandarin words to be used, it was important to consider their difficulty for novel learners. This is particularly pertinent, as when the phonetic demands are high during L2 learning, gestures can in fact hinder learning (Kelly & Lee, 2012).

A total of 24 Mandarin (Pinyin) action words were used in this study (see Appendix B, Table B1, p.165, for the full word list used). These verbs were chosen as they could be demonstrated easily as both simple actions using an object, as well as a re-enactment without the object present (as a gesture). The action words selected were simple, everyday actions such as pour, whisk and hammer. These words were selected based on previous studies on co-speech gestures (Kelly et al., 2010; Zhao et al., 2018) and from the action words used in a the pretest validation studies described above (see Appendix C, p.167, for semantic congruency effect by word pairs). Of the 44 actions words used in the pretests, 24 final words were chosen for this study.

The final words were selected with the aim of minimizing phonological and semantic similarity across the words and gestures. In total, 18 of the words were single toned words and 6 were two-toned words. The two-toned words were split evenly across the two learning condition word sets (A and B) that were created. A native Mandarin speaker helped with the decision on the final word list and advised which words they believed would sound and appear too similar to a beginner learner. To this aim, no two words were selected that only differed in lexical tone. Additionally, words that only differed by an unfamiliar phoneme, or consonants that would only be distinguishable if they were voiced/unvoiced were avoided. Out of the 24-word set, there are only 3 words that rhyme (sǎo, dào and yáo) but these vary by tone and by their onset consonant that is very close to the equivalent in English. Semantic overlap between the gestures was reduced by minimizing the number of gestures that had similar movement shapes in the stimulus set. By doing so, the final stimuli set did not include any words or gestures that were too alike.

The study also used 48 action video clips (24 with a male actor and 24 with a female actor) as stimuli. In these clips, the actions were performed using the appropriate items and equipment. Additionally, 48 gesture video clips (24 male and 24 female) were used as stimuli in the study. In the clips, the 24 action words were performed using only gestures. When performing the actions and gestures in the video clips, the actors were sat at a table with a plain background and always said the action word whilst carrying out the movement (see Figure 7). When recording the video clips, the actors first recorded the action clip for a word and then recorded the gesture. Different gesture ideas were discussed by the actors and researchers.

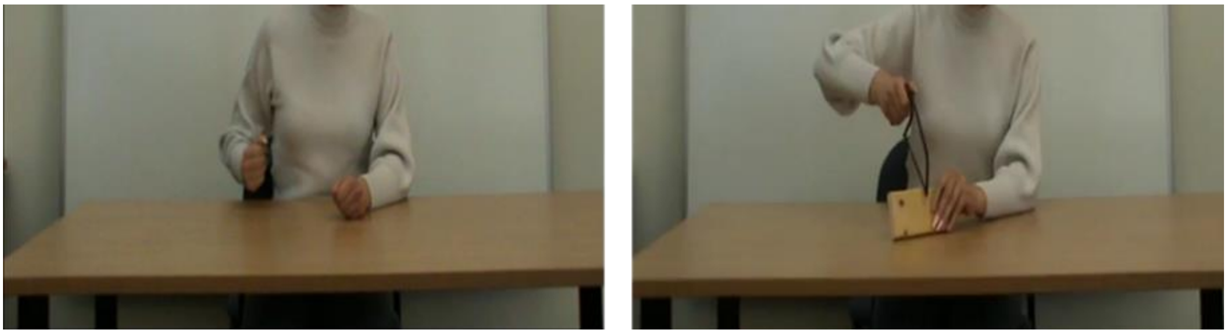


Figure 7. Freeze frames from video clips showing the action word 'saw' being performed as a gesture (left) and as an action (right). As described above, gesture clips were used during the learning phase, whereas action clips were used in the assessment phase.

All of the video clips were edited and cropped using OpenShot software so that the heads and necks of the actors were not in view. The action videos were also cropped so that they all started with the first stroke of the action. This was necessary for the action clips as these were used as the video stimuli in the implicit reaction time assessment task and so needed to all start with the onset of the action to ensure that all actions would begin at the same time. The action videos were cropped into short clips no longer than 2000 milliseconds (ms) long. Due to the nature of the different actions, the clips did vary in length. The longest action video clip was 2000ms and the shortest was 450ms (M = 1082ms, SD = 458ms).

The gesture videos were also cropped into shorter clips but were not edited to start with the first stroke of the gesture. This was because the gesture clips were used in the learning phase of the experiment and so the gesture needed to be as clear as possible. The longest gesture clip was 2390ms and the shortest was 1020ms (M = 1536ms, SD = 402ms).

Audio stimuli of the action words were recorded separately in both English and Mandarin by both a male and female speaker to improve the sound quality. Recordings took place in a sound-proof lab on the University of Hull campus in order to eliminate any background noise from the recordings. These high-quality recordings were used in all learning and assessment procedures instead of the low-quality audio from the video recordings.

Procedure

The different sections of the online study were all developed using Psychopy software (Peirce et al., 2019). The sections were synced onto Pavlovia where it was licenced by the University of Hull. This enabled the experiment to be shared via a URL, run online and the data to be collected. An online survey was also created, using the website Jisc, to provide participants with information, to gain consent and to collect a few demographic details (such as gender, age and the languages spoken by them) from the participants at the beginning of the experiment. The survey also asked about participants' motivation to learn Chinese and was used to check that no participants had any previous knowledge of the Mandarin words that they would be learning. At the end of the short survey, it redirected them to the first part of the study on Pavlovia.

As participants signed up to the study, they were randomly assigned to either group A or B. This counterbalanced which Mandarin words were taught using gestures and which were taught using English translations. The study consisted of three experimental sessions, a learning session and then two assessment sessions, with the first assessment taking place one day after learning, the second assessment taking place seven days after learning (see Figure 8 for a timeline of the sessions). All of these sessions took place online, with participants using their own personal computer and laptops to complete them.

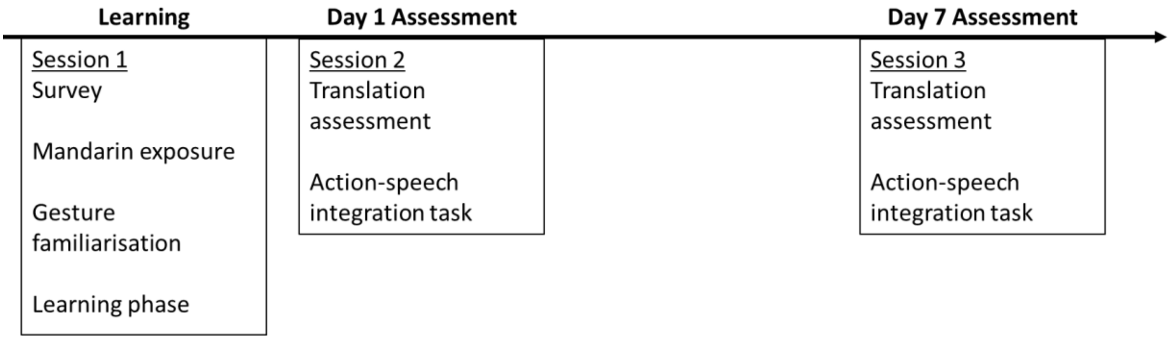


Figure 8. The timeline for all sessions in the experiment.

Learning session

In the first session, a short survey was used to gain online consent, to collect basic demographic information from the participants and to ensure that participants did not have any previous knowledge of the Mandarin vocabulary that was being used.

Afterwards, participants were redirected to a Mandarin exposure phase. This was included to give participants the chance to become familiar with the novel phonetics and script of Mandarin words. In this phase, participants were exposed to all 24 Mandarin words before any learning took place. For each word, the pinyin was presented in the centre of the screen. Participants then heard each Mandarin word being said twice by the female teacher and then twice by the male teacher. The pinyin remained on the screen and after 10 seconds a message appeared that reminded participants to press the space bar to continue to the next word. The order of the Mandarin words was randomised.

In the subsequent familiarisation phase, participants viewed the 12 gestures that they would be using to learn half of the Mandarin words. This familiarisation phase was included to minimise any ambiguity that the gestures might have, by informing participants of their exact meaning. The gestures viewed depended on whether the participant was in group A or B. Participants viewed all of the gestures from one teacher (either male or female) before watching all of the gestures from the other teacher. The order that participants saw the teachers in was randomised across participants. In each gesture exposure trial, the short gesture video clip was played twice. At the same time, the matching gender speech was played that named the gesture in English, twice also. Participants were reminded to press the spacebar to continue to the next gesture. Once all the gesture trials from the first gender were shown twice through, the gestures from the other teacher were shown in the same way.

Next, participants were redirected to the learning phase. The 24 Mandarin words that the participants were taught were split into the 12 to be learned through gestures and the other 12 to be learned through English translations. The learning method used first was randomised across participants. Learning was broken down into 3 blocks of 4 words for each learning condition (see Appendix B, Table B2, p.165). In each block, the four words were presented and taught first by the female teacher and then by the male teacher. Participants were presented with each word on the screen in pinyin text, at the same time the matching Mandarin speech was presented audibly. In each learning trial, the pinyin and Mandarin speech were presented twice followed by the presentation of a learning cue. For half of the words, the learning cue consisted of an English translation in text form, for the other half of the words, it consisted of a gesture clip (performed by the teacher of the same gender as the speech; see Figure 9). Whilst the learning cue was displayed on the screen, the participant was presented with the Mandarin speech again. Each learning cue was presented twice for each word (to view an example block, see ResearchBox, Study 1, 'Materials', https://researchbox.org/2841). In total, within each block, participants were exposed to the learning cue of each word 4 times (twice with the female teacher and then twice with the male teacher).

Figure 9. Freeze frames from the learning phase displaying the translation learning cue (left) and the gesture learning cue (right).

After each learning block, participants completed a mini test which involved multiple choice translation tasks on the 4 words they had been presented with in the previous block. This was used to assess whether learning was taking place, and to check that participants were not simply pressing buttons ad-hoc. The mini test also helped to keep participants engaged in the learning process. In this task, the translation/gesture that corresponded to one of the Mandarin words learnt in the previous block would be presented on the screen with the text '=?' alongside it. Then the pinyin of the 4 Mandarin words would appear, one by one, from left to right, along the bottom of the screen. As each word appeared, the audio for the word was played once. Participants were instructed to press numbers 1 to 4 on their keyboards to indicate the correct Mandarin word that corresponded with the translation/gesture. Once a number was selected, a message appeared that stated either 'Correct!' or 'Incorrect!' along with the correct Mandarin word. This mini test was carried out for each of the 4 words in the block, first with the female teacher and speech and then with the male teacher and speech. Once every block and mini test had been completed, the process was repeated in the same order (to view an example mini-test, see ResearchBox, Study 1, 'Materials', https://researchbox.org/2841). Thus, in total, participants completed the 6 blocks twice through, resulting in 96 mini-test trials.

Day 1 and Day 7 Assessment sessions

The next day, participants completed the second session. Participants were told that ideally this session should be completed at around the same time as the previous day (24 hours after learning). However, some flexibility was allowed in order to fit around participants' schedules and to encourage full participation. The first assessment session (day 1) was completed, on average, 24.63 hours (*SD* = 4.95 hours, *min* = 13.98 hours, *max* = 46.72 hours) after learning took place.

Due to technical problems, two participants completed the translation task for the first assessment session on the correct day, but then completed the action-speech integration task 1 day late (2 days after learning). Two other participants completed the first assessment sessions (translation and action-speech integration task) 2 days after learning took place (31.12 hours and 46.72 hours). These participants remained in the data set. As mentioned before, one participant that completed the first assessment session 3 days after learning took place (64.43 hours) was already removed from the data set.

In each assessment session, participants first completed the explicit translation task described above. Once all 24 words had been presented, participants were redirected to the action-speech integration. The task consisted of 3 blocks, with breaks in between the blocks. Each block contained 64 individual video clip trials, with block order randomised across participants. The order of the individual videos within each block was also random. When participants reached a break after completing a block, they were told their accuracy percentage across the trials in that block as well as their average reaction time. Participants were reminded of each of the keys they were required to press for each gender before starting the next block. This action-speech integration task took participants around 20 minutes to complete on average.

The whole assessment phase was also completed again 1 week after the learning phase took place. This second assessment session (day 7) was completed, on average, 172.36 hours after learning (*SD* = 9.05 hours, *min* = 160.11 hours, *max* = 203.04 hours). A total of six participants completed this assessment one day late (*Mean hours after learning* = 192.19). These participants remained in the data set.

Transparency and Openness

This study was pre-registered online using AsPredicted. A copy can be found in the ResearchBox folder for this thesis (https://researchbox.org/2841) under the section 'Study 1'. *Deviations from pre-registered protocol:* Only participants with a full data set were used in analysis as results were analysed using a 2x2x2x2 ANOVA using day as a factor and not analysed separately for each day.

Data analysis

<u>Mini-test performance</u>

Mini-test data from the 54 participants that completed the full learning phase was gathered. Overall, for the 96 mini-tests involved during learning, participants had an accuracy of over 90% (*M* = 87.85, *SD* = 13.37). Four participants were identified as having mini-test scores 2 *SD*s, or more, below the mean (<61.12) (see Appendix D, table D1, p.168, for Study 1 mini-test data). Of these four participants, one only completed the learning phase. Therefore, of the 46 participants that completed all three sessions of the study, 43 remained for analysis.

<u>Action-speech integration task outliers</u>

Outliers for the action-speech integration task were filtered out. All trials less than 200ms were removed as these responses were considered particularly fast, pre-emptive trials (Luce, 1986, Whelan, 2008). Additionally, trials with reaction times over 2000ms were also removed as these would have been 'timed-out' if the study had been conducted in a laboratory setting. Additionally, all incorrect trials were also removed.

Lastly, participant outliers were removed, these were classed as trials that were plus or minus 2.5 *SD*s from participants mean reaction times (as recommended by Ratcliff, 1993). The data set was then checked to see if any participants had less than 80% of trials left after outlier rejection. In total, three participants had less than 80% of trials remaining (one participant had less than 80% on both days, two participants had less than 80% for day 7 assessment). These participants were removed from the whole data set, leaving 40 participants for analysis.

For day 1, the participant with the lowest number of trials remaining still had 85.42% of the original data set remaining (*M* = 92.62%, *SD* = 3.22%). For day 7, the smallest percentage of trials remaining for a participant was 80.73% (*M* = 92.34%, *SD* = 4.02%).

Results

Translation

In the translation assessment task, answers were scored as correct if they were the exact translations given in either the gesture familiarisation phase (for words learnt through gestures) or in the learning phase (for words learnt through translations). Additionally, conjugations (e.g., past participles) and direct synonyms were also accepted as correct responses.

For the first assessment task, one day after learning, on average the 43 participants correctly translated just over half of the Mandarin words (*Mean percentage accuracy* = 54.17%, *SD* = 27.26%). More words were correctly translated from the gesture learning condition (*M* = 58.14%, *SD* = 29.35%) than from the translation learning condition (*M* = 50.19%, *SD* = 28.67%).

For the second assessment task, 7 days after learning, the participants correctly translated just under half of the Mandarin words (*M* = 49.90%, *SD* = 25.73%). More words were correctly translated from the gesture learning condition (*M* = 54.46%, *SD* = 27.66%) than from the translation learning condition (*M* = 45.35%, *SD* = 27.84%).

A 2-way ANOVA was conducted using the factors day and learning method (see Figure 10). There was a significant effect of learning method, $F(1,42) = 9.09$, $p = .004$, $\eta^2_g = .023$, indicating that more words were correctly translated from the gesture condition (*Mean accuracy* = 0.563, *SE* = 0.0407) than from the translation condition (*M* = 0.478, *SE* = 0.0404). There was no significant effect of day, $F(1,42) = 2.43$, $p = .127$, and no interaction between learning method and day, $F(1,42) = 0.21$, $p = .649$. Post-hoc tests indicated the effect of learning method was significant both at the first (*M* = 7.95%, *SE* = 3.10%, $p = .0104$) as well as at the second assessment day (*M* = 9.11%, *SE* = 3.10%, $p = .0033$).

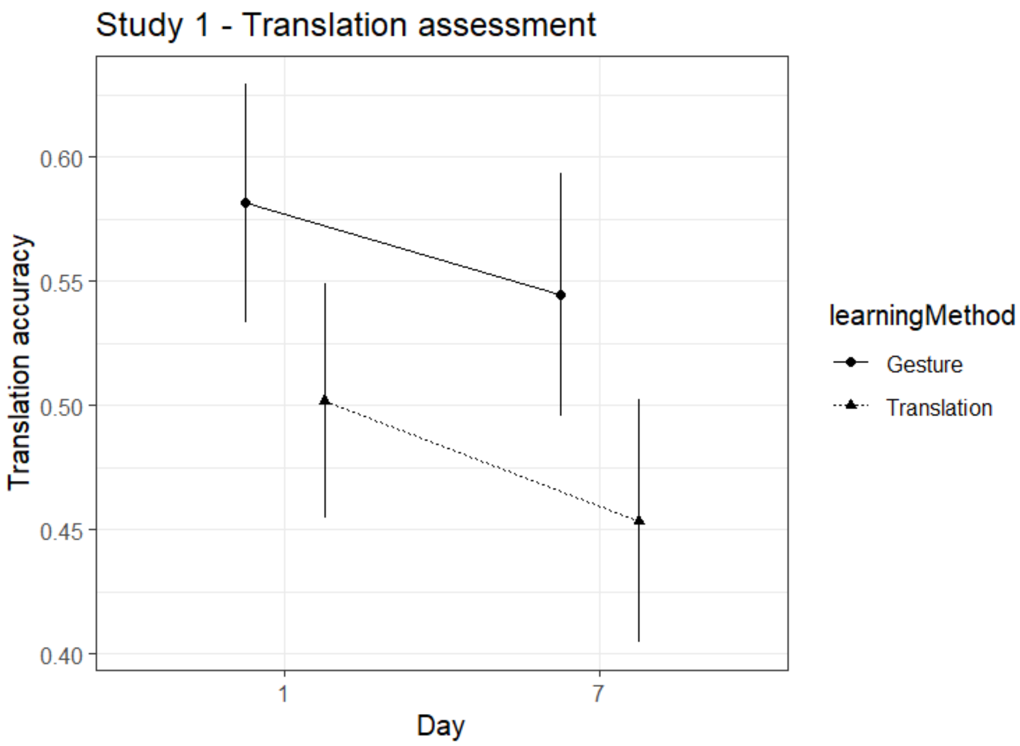


Figure 10. Graph displaying the mean translation scores across assessment day 1 and 7 for words learnt through gesture cues and translation cues. The within-subject error bars reflect the variability of the data using 95% confidence intervals.

Action-speech integration task

The reaction time data was analysed using a 2 (learning method) x 2 (day) x 2 (semantic congruency) x 2 (gender congruency) ANOVA (see Figure 11). There was a significant main effect of gender congruency, $F(1,39) = 5.91$, $p = .020$, $\eta^2_p = .13$, indicating longer reaction times for gender incongruent (*M* = 687ms, *SE* = 26.9ms) than congruent trials (*M* = 679ms, *SE* = 28.1ms). A significant two-way interaction was observed between learning method and semantic congruency $F(1,39) = 4.76$, $p = .035$, $\eta^2_p = .11$, indicating that the size and direction of the semantic congruency effect was affected by learning method. For words learned via gesture, semantically incongruent trials produced longer reaction times (*M* = 688, *SE* = 27.3) than semantically congruent trials (*M* = 679, *SE* = 27.0). A reversed pattern was observed for words learned via translation, where shorter reaction times were observed for semantically incongruent (*M* = 680 *SE* = 27.8) as compared to congruent trials (*M* = 685, *SE* = 28.1). Additionally, there was a significant interaction between learning method and day, $F(1, 39) = 5.50$, $p = .024$, $\eta^2_p = .12$, indicating that the decrease in reaction times from day 1 to day 7 was more pronounced for words learned via translation (*M day 1* = 698, *SE* = 31.3; *M day 7* = 666, *SE* = 26.8) than for words learned via gesture (*M day 1* = 693, *SE* = 30.3; *M day 7* = 674, *SE* = 26.2). All other main effects or interactions were not significant.

Although I did not observe a three-way interaction of day × learning method × semantic congruency, I additionally tested whether there was an interaction between learning method and semantic congruency within each assessment day, to see whether it is robust within each assessment occasion. For the day 1 assessment, there was a significant interaction, $F(1,39) = 5.56$, $p = .023$. There was no significant interaction between semantic congruency and learning method for the day 7 assessment, $F(1,39) = 0.83$, $p = .368$.

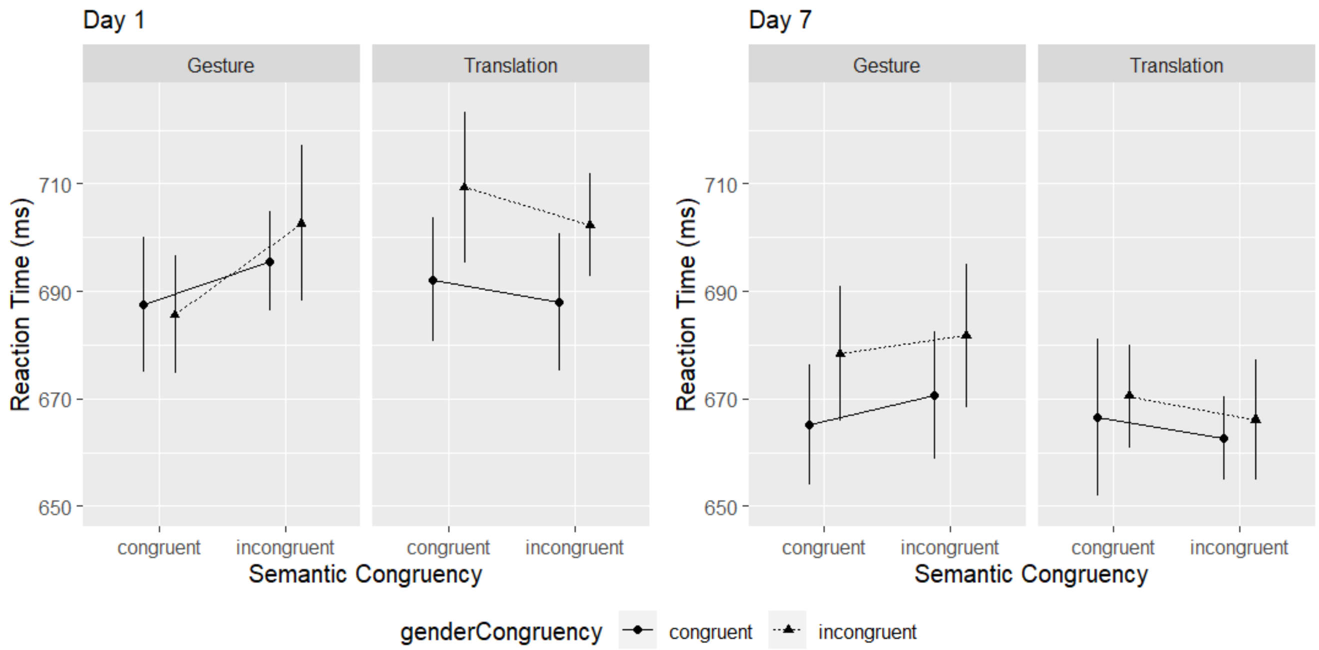


Figure 11. *Graph displaying the reaction times for the Action-speech integration assessment across the factors day, learning method, semantic congruency and gender congruency. Within-subject error bars reflect the 95% confidence intervals.*

Discussion

Study 1 had three clear aims: to determine if the gestural advantage during L2 word learning would remain evident under speeded implicit assessment conditions; to investigate if gesture would still be a beneficial cue to word learning when presented as the only cue to the L2 word; and to examine the retention of the gestural advantage over time and after several consolidation periods. The main findings of this study were that gestures, when presented on their own, had greater translation accuracy in the explicit assessment over L1 translation cues, with this advantage persisting 7 days after learning. Additionally, in the action speech integration task, semantic congruency effects were evident for words taught through gestures but not L1 translations, however this effect was not found on day 7.

Translation task

In the explicit L2 to L1 translation task, significantly more words were translated from the gesture learning condition than from the L1 translation learning condition. This effect of learning method was persistent over time, with gestures still producing greater translation scores a week after learning. This supports the previous literature (Macedonia & Von Kriegstein, 2012) showing a robust gesture advantage. Compared to the past research, this study demonstrated that this gestural benefit still remained despite gestures being used as the only cue to the L2 word. This extends Tellier's (2008) research, indicating that the effect is also consistent in adult learners with a larger word learning set.

It is also notable that a significant effect of gesture cues was found in this assessment task, considering the lack of exposure to the L1 translation word in this learning condition. The translation task required the L1 translation to be given in response to the L2 word, with gestures not included in the task. Such a task could be quite difficult for L2 words that have been learnt solely through gestures. Furthermore, this task could be argued to be more favourable to words learnt through the translation condition as they had received plenty of exposure to the L1 words. Additionally, the assessment task was more similar to the learning session which could also be advantageous. As this study used gestures presented in isolation, without accompanying L1 translations, the assessment task conditions were very different to the learning environment.

These findings could have considerable implications to the teaching and pedagogy of second language learning. As gestures do not require the accompanying L1 to be beneficial, this information may be omitted during learning without any loss in the effectiveness of learning. This has the potential to allow for a quicker, more streamlined and L2 focussed learning environment and approach.

Action-speech integration task

Study 1 also found evidence for a gestural advantage when learning was tested implicitly through the action-speech integration task. Words learnt through gestures displayed semantic congruency effects, with incongruent stimuli resulting in longer reaction times than congruent stimuli. However, the same effect was not evident for words learnt through L1 translations. These results indicate that L2 words learnt through gestures had been semantically integrated into the mental lexicon as semantically incongruent stimuli showed interference effects (longer reaction times than congruent stimuli). However, the same integration was not evident in words learnt through L1 translations, with no semantic interference effects being displayed (i.e. responses did not differ if the stimuli was congruent or incongruent).

This extends the past literature by demonstrating that the gestural advantage is not limited to simple, explicit, controlled retrieval tasks. Gesture cues were able to aid the deeper semantic understanding of L2 words to such an extent that they displayed automatic semantic integration. Interestingly, the effect of learning method on semantic congruency was not evident in the assessment one week after learning. This suggests that, unlike the explicit knowledge, the semantic connections measured in the implicit assessment may be too fragile and underdeveloped in these early stages of learning to remain present over an extended course of time.

Understanding the gesture advantage

The gestural advantage previously reported in the literature was found to persist in this study when gesture was used as the only cue to L2 word learning and in a speeded, implicit task. There are several possibilities for this advantage that need to be explored.

Firstly, as discussed in the introduction to Study 1, the RHM could be used to explain this advantage, particularly in relation to the implicit assessment. By including only iconic gestures that hold semantic information during learning, these cues could be developing the direct conceptual links between L2 and meaning that usually only become developed in advanced L2 learners. By limiting exposure to the L1 during learning (having only been available during initial introduction to the gestures during a familiarisation phase) the typical lexical route was not activated. Due to the development of the direct conceptual links with the words learnt via gestures, these words would then have displayed more semantic integration during the implicit task.

Another explanation for the advantage could come from the afore mentioned privileged access account. When compared to simple L1 translations in L2 acquisition, gestures have access to specialist motor traces and representations which create more memorable learning for the L2 words. However, Study 1 was not able to test this mechanism in isolation as disambiguation could also be involved in effects of gesture-based learning.

Another possible mechanism driving the gesture advantage in L2 learning could be disambiguation. When considering the stimuli used in the study, the gestures involved during learning provided specific details about the context of the action words. For example, the gesture depicts opening a book, as opposed to opening a door or a bag of crisps. In contrast, the translation learning condition was semantically underspecified, only providing the English word 'open'. The gestures had the ability to eliminate other applications of the words, allowing learners to focus on one specific example, whereas in the translation learning condition no such contextual example was provided. Could the gesture advantage be driven, at least somewhat, by this extra context provided, that in turn creates a more memorable learning environment? This is a question that must be investigated further.

To conclude, the gestural advantage was found to persist when used in isolation and in both explicit and implicit tasks. This advantage was robust over time in the translation task but not in the implicit action-speech integration task, suggesting that the explicit effects may be more durable. The effective use of gestures as the sole cue in L2 vocabulary learning supports the RHM, although it is still unclear from this study alone exactly what mechanism drives this advantage.

STUDY 2 – THE ROLE OF DISAMBIGUATING CONTEXT ON THE GESTURAL ADVANTAGE

Introduction

Study 1 revealed the persistence of the gestural advantage in a speeded implicit assessment and when gestures are presented on their own. This extended the previous literature which often contained gestures alongside other cues as well as measuring memory for L2 words using explicit behavioural measures (e.g., Porter, 2012; Tellier, 2008). However, Study 1 did not explore the potential mechanisms driving this advantage. Study 2 will aim to investigate one of the mechanisms that could be underlining the effect, the disambiguation account.

Iconic gestures have been found to be a useful cue for disambiguating speech when homophones are used (Holle & Gunter, 2007). Additionally, gestures are attended to more in suboptimal listening conditions (Rogers, 1978). Obermeier et al. (2012) found that when the communicative environment was overloaded with babble noise, gestural information had greater integrated with speech than in a noise-free setting. This demonstrates the beneficial disambiguating qualities of gestures. The disambiguation account suggests that the contextual nature of iconic gestures can aid learning by creating a more defined learning environment. This additional disambiguation provided is even more apparent when gestures are compared to simple L1 translations. In Study 1, when comparing the two learning methods used, the gesture learning condition provided more information on the specific use of the verb than the translation condition did. In the gesture condition, the videos showed the context of the verb, for example zipping up a jacket rather than a bag or typing on a keyboard and not on a phone. It may be the case that having this additional information provided by gesture creates a more memorable learning environment.

Additionally, the gesture videos gave participants clarification on some of the words that had more than one meaning, for example the Mandarin word for saw was presented with a gesture video of the actor sawing a piece of wood, rather than looking and seeing something. This begs the question whether it is the use of gestures specifically that leads to this advantage in L2 word learning, or whether the additional information about the new words that they provide makes them more memorable. In particular, this information on the context of the actions could be the reason for the better learning outcomes for the gesture condition as compared to the translation condition observed Study 1.

Thus, in order to test this disambiguation account, this next study aimed to test whether this advantage of gestures is still present when the translation condition is made unambiguous, by including the same information that is provided through the gestures. Study 2 therefore compared the use of gesture cues to translation plus example text cues. The additional example text (e.g., to zip up a jacket) matched the semantic information presented through gesture. If the effect is somewhat driven by disambiguation, then the gestural advantage should reduce or disappear given the equal amount of contextual information now provided across the two learning conditions. It is important to note that this study was not able to test disambiguation in isolation. As the two learning conditions still differed in the inclusion of a gesture learning cue, another mechanism, such as the privileged access account, may also be involved. Rather, Study 2 assesses if disambiguation is a factor that could be involved in the positive effect of gestures in learning.

Finally, Study 2 will also test if the effects in the action-speech integration task found in Study 1 are replicable. In Study 1, the semantic congruency effects found in the implicit assessment were only evident one day after learning, as opposed to the translation accuracy learning effects which also presented 7 days after learning. Considering this, the strength of these effects is questioned. Additionally, since this is the first time this implicit action-speech integration task has been used to assess automatic semantic integration of newly learnt L2 words, it is important that a reliable effect can be found.

Aims for Study 2

To sum up, this study aimed to investigate the effect of contextual information on L2 word learning. The study explored if providing the same contextual information that is supplied from gestures to an L1 translation condition could reduce the gestural advantage. The following questions are investigated:

i. Can the gestural advantage be explained, at least in part, by the disambiguation account?
ii. Does the effect of gesture on the implicit speeded task observed in Study 1 replicate?

Method

Participants

Recruitment for this study was completed using both the University of Hull's SONA participant recruitment system and Prolific. Participants that signed up via the University of Hull's SONA website and were paid with 2.5 research course credits and a £10 Amazon voucher. The voucher reward was only awarded for completion of the full study. Participants that were recruited through Prolific were paid as close to £8/hr as possible (payment was set to £8.12/hr for session 1 and £8/hr for session 2 and 3). These payments were set to the predicted average time to complete and may have fluctuated slightly based on the participant sample median completion time. The study was only visible to participants within the UK and with a minimum of 10 previous submissions. The study was not visible to those fluent in Mandarin, Cantonese or Chinese. In total, 21 participants were recruited from Prolific and 64 from SONA. Data collection took place from February to April 2022.

In total, 85 participants enrolled in the study, however one participant withdrew from the study, seven did not complete the first learning stage, one completed the learning phase twice and so was removed, seven completed only the learning stage and no assessment and nine did not complete the final assessment session. Additionally, one participant completed the final session too late. Participants were given some leeway with completing the assessment sessions, with the maximum time allowed being set to one day late (2 days after learning for the day 1 assessment and 8 days after learning for the day 7 assessment). This participant completed the session outside of this permitted time period (9 days after learning took place) and was therefore removed from the analysis. One participant was also removed for completing the first assessment session too early. They completed the learning phase at gone midnight on one day and then completed the first assessment before midday (only 10.34 hours between the two sessions).

The remaining participant group of 58 that completed the study contained 32 females and 26 males. The mean age of participants was 27.69 (SD = 10.91). Of these participants, 40 were recruited from SONA and received course credit and the voucher and 18 were recruited from Prolific and received payment averaging a rate of £8/hr. For 55 of the participants, English was their first language. For the other three participants, all stated English as a fluent second language. All of the participants except one reported that they had no previous knowledge of any Chinese vocabulary. The one participant that answered yes, did not know any of the translations for any of the action words used in the study. In a self-reported question on personal motivation to learn Chinese, 35 participants (60.34%) stated that they were personally motivated.

Design

The same within-subject design used in Study 1 was used in this second study. The translation task used a 2 x 2 factorial design, using the two factors Learning Method (gesture, translation plus example) and Day (1, 7). For the action-speech integration task, a 2 (learning method) x 2 (day) x 2 (semantic congruency) x 2 (gender congruency) design was used.

Participants first completed the learning phase which included the gesture familiarisation phase, the Mandarin exposure phase and the learning session. Within this session, participants learnt 12 of the Mandarin words through gestures and the other 12 words through translation plus example text. This session was designed in the same way as Study 1, with a total of 12 blocks and 96 mini-tests. The assessment sessions that took place 1 day and 7 days after learning were also the same as that used in Study 1, with both the explicit translation assessment and the implicit action-speech integration task.

Stimuli

The same 24 Mandarin words were used for this study, as well as the same gesture video clips for learning and action video clips for assessment. Example text was created to accompany the English translation words. This text was created to provide the same information on the context of the words as the gesture videos (see Table 2 for full list of example text).

Table 2

List of the additional text created to accompany the English translation during the learning phase.

Mandarin word (Pinyin)	English translation	Additonal example text
guān	close	(e.g. to close a book)
jù	saw	(e.g. to saw a piece of wood)
sǎ	sprinkle	(e.g. to sprinkle some glitter)
diǎn rán	light	(e.g. to light a match)
gǎn	roll	(e.g. to roll a rolling pin)
chuí	hammer	(e.g. to hammer a nail)
jiǎn	cut	(e.g. to cut paper with scissors)
dǎ zì	type	(e.g. to type on a keyboard)
jié	knot	(e.g. to knot a rope)
mǒ	spread	(e.g. to spread butter on bread)
dào	pour	(e.g. to pour a drink)
kāi suǒ	unlock	(e.g. to unlock a padlock)
pēn	spray	(e.g. to spray deoderant)
kāi	open	(e.g. to open a book)
jiǎo bàn	whisk	(e.g. to whisk a mixture in a bowl)
sǎo	sweep	(e.g. to sweep a table)
rēng	throw	(e.g. to throw a ball)
jǐ	squeeze	(e.g. to squeeze a sponge)
nòng duàn	snap	(e.g. to snap a stick)
xiě	write	(e.g. to write on paper)
sī	tear	(e.g. to tear a piece of paper)
bō	dial	(e.g. to dial a number on a phone)
yáo	shake	(e.g. to shake a bottle)
lā shàng	zip up	(e.g. to zip up a jacket)

A short survey was created to gain additional information on the data collection after all sessions had been completed. Questions asked participants about their behaviour in the learning phase (whether they repeated the words out loud, whether they performed the gestures themselves etc.) and in the assessment phase (how they responded if they were unsure in the translation task, whether they always looked at the screen during the action-speech integration task). However, the data collected was qualitative and so was not used during analysis of this study. The results were used to guide the development of a quantitative end of study survey used in Study 4 and 5.

Procedure

The procedure for this study was the same as that in Study 1. If participants had signed up via SONA, the link to the first session was sent by 9am on the morning of their booked day. Participants were randomly allocated to either gesture set A or B which determined which set of 12 Mandarin words were taught through gestures. If participants signed up via Prolific, they began the session on the same day that they signed up. Two studies for session 1 were created on Prolific for gesture set A and B.

Learning session

During the learning session, participants completed the Mandarin exposure phase, the gesture familiarisation phase and then the learning phase. During the learning phase, participants learnt 12 of the Mandarin words through gestures and 12 through translations plus examples (see Figure 12). Example text was displayed underneath the English translation text.

Figure 12. Freeze frames from the learning phase of the Mandarin word "mǒ" displaying the translation plus example learning cue (left) and the gesture learning cue (right).

Day 1 and Day 7 Assessment sessions

For the second session, participants recruited through SONA were sent their session link by 9am the day after learning was completed. Prolific participants were invited to a 'Session 2 Study' by 9am on the day after learning. When added, participants received an email notification.

The first assessment session (day 1) was completed, on average, 23.47 hours (*SD* = 5.12 hours, *min* = 13.34 hours, *max* = 40.37 hours) after learning took place. One participant completed the first assessment session 2 days after learning took place (40.37 hours). This participant remained in the data set. As mentioned previously, there was one participant that completed the day 7 assessment session 10 days after learning took place (228.79 hours) and one participant that took the first assessment session too early (10.34 hours). These two participants were already removed from the data set and so were not included in these descriptive statistics.

The same process was used for the third session for participants that had completed session 2. Seven days after learning was completed, participants were added to the 'Session 3 Study' on Prolific or sent the session link. This day 7 assessment was completed, on average, 166.91 hours (*SD* = 5.03 hours, *min* = 158.59 hours, *max* = 186.12 hours) after learning took place. One participant completed this day 7 assessment 1 day late (186.12 hours). This participant remained in the data set.

After session 3, participants had a short end of study survey to complete. For SONA participants, this survey was emailed to each participant upon completion of the final assessment session. For Prolific participants, after the action-speech integration task on day 7, they were automatically redirected to the end of study survey.

Transparency and Openness

This study was pre-registered online using AsPredicted. A copy can be found in the ResearchBox folder for this thesis (https://researchbox.org/2841) under the section 'Study 2'. Note that the order in which the studies are being presented in this text is different to the chronological order in which the studies were conducted. Therefore, on the pre-registration form, this study has the title 'Study 3' instead of 2. *Deviations from pre-registered protocol:* Only participants with a full data set were used in analysis as results were analysed using a 2x2x2x2 ANOVA using day as a factor and not analysed separately for each day.

Data Analysis

Mini-test performance

Mini-test data from the 76 participants that completed the full learning phase was gathered. Overall, for the 96 mini-tests involved during learning, participants had an accuracy of over 90% (*M* = 88.58, *SD* = 7.49). Five participants were identified as having mini-test scores 2 *SD*s, or more, below the mean (<73.60) (see Appendix D, table D2, p.168 for Study 2 mini-test data). Of these five participants, one only completed the learning phase, and another did not complete the day 7 assessment. Therefore, of the 58 participants that completed the study, 55 remained for analysis (32 female*, M age* = 27.55, *SD* = 10.98).

Action-speech integration task outliers

Outliers for the action-speech integration task were calculated. First, the accuracy, mean reaction times and standard deviations of the 55 participants remaining for analysis were checked. When overall accuracy across both days was checked, seven participants had accuracy lower than 80% (73.96%-79.17%). After these were removed, accuracy scores when split by day were reviewed. Two further participants had accuracy lower than 80% for day 1 (77.08% and 78.65%) and so were removed from the analysis.

Outliers were then filtered out, with all trials less than 200ms being removed as pre-emptive trials, all reaction times over 2000ms being removed as 'timed-out' trials and all incorrect trials also being removed. Finally, individual participant outliers were removed (trials +/- 2.5 *SD*s from participants mean reaction times).

After these rejections, a further six participants had less than 80% of trials remaining (three participants had too few trials on both days, one participant had too few for day 1 assessment and two participants had too few remaining for day 7). These participants were also removed from the whole data set. In total, 15 outliers were removed from the data set, leaving 40 participants for analysis in the action-speech integration task.

For day 1, the participant with the lowest number of trials remaining still had 84.90% of the original data set remaining (*M* = 93.20%, *SD* = 2.88%). For day 7, the smallest percentage of trials remaining for a participant was 80.73% (*M* = 91.46%, *SD* = 4.03%).

Results

Translation

For the first assessment task, one day after learning, on average the 55 participants correctly translated over half of the Mandarin words (*Mean percentage accuracy* = 59.17%, *SD* = 26.03%). More words were correctly translated from the gesture learning condition (*M* = 59.85%, *SD* = 26.65%) than from the translation plus example learning condition (*M* = 58.48%, *SD* = 28.62%).

For the second assessment task, 7 days after learning, the participants correctly translated just over half of the Mandarin words (*M* = 52.73%, *SD* = 29.47%). The number of correctly translated words were very similar for the gesture learning condition (*M* = 53.18%, *SD* = 31.20%) and the translation plus example learning condition (*M* = 52.27%, *SD* = 31.03%).

A 2-way ANOVA was conducted using the factors day and learning method (see Figure 13). There was a significant effect of day, $F(1,54) = 17.15$, $p < .001$, $\eta^2_g = .012$, indicating more correct responses on day 1 (*Mean accuracy* = 0.592, *SE* = 0.0351) than on day 7 (*M* = 0.527, *SE* = 0.0397). There was no significant effect of learning method, $F(1,54)$

= 0.24, p = .624, and no interaction between learning method and day $F(1,54) = 0.03$, p = .854.

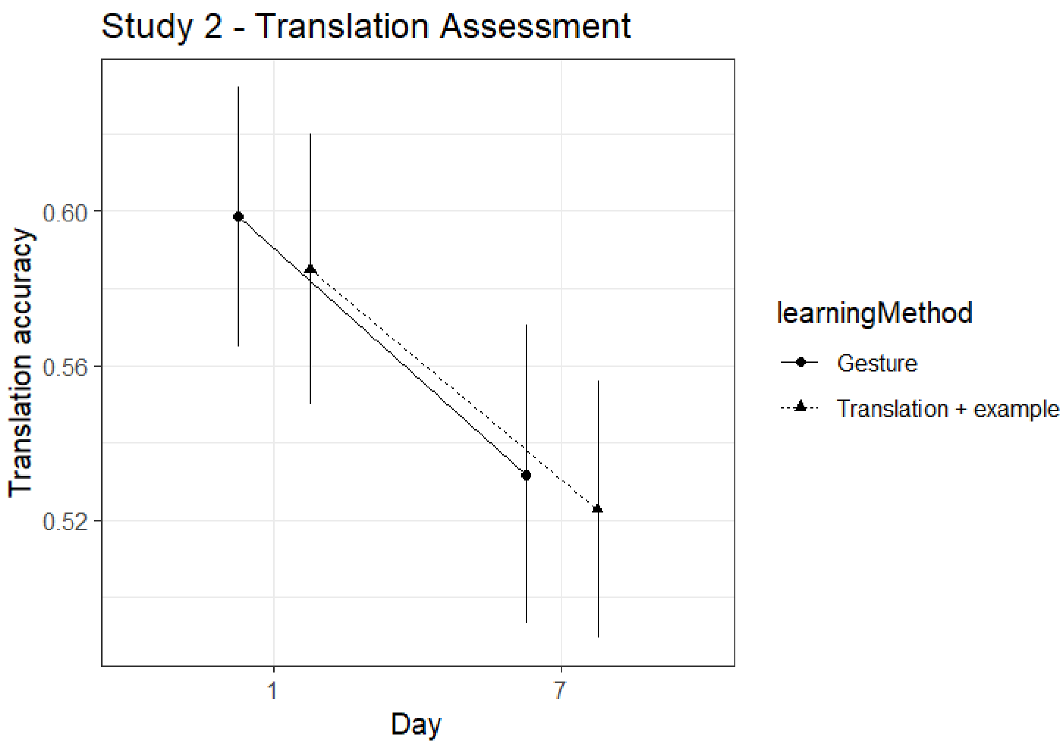


Figure 13. Graph displaying the mean translation scores across assessment day 1 and 7 for words learnt through gesture cues and translation plus example cues. The within-subject error bars reflect the variability of the data using 95% confidence intervals.

Action-speech integration task

The reaction time data was analysed using a 2 (learning method) x 2 (day) x 2 (semantic congruency) x 2 (gender congruency) ANOVA (see Figure 14). There was a significant main effect of gender congruency, $F(1,39) = 18.27$, $p < .001$, $\eta^2_p = .32$, indicating longer reaction times for gender incongruent (M = 672ms, SE = 23.8ms) than congruent trials (M = 659ms, SE = 24.1ms). There was no significant effect of semantic congruency, $F(1,39) = 2.64$, p = .112, or interaction between learning method and semantic congruency, $F(1,39) = 0.00$, p = .958. All other main effects and interaction were not significant.

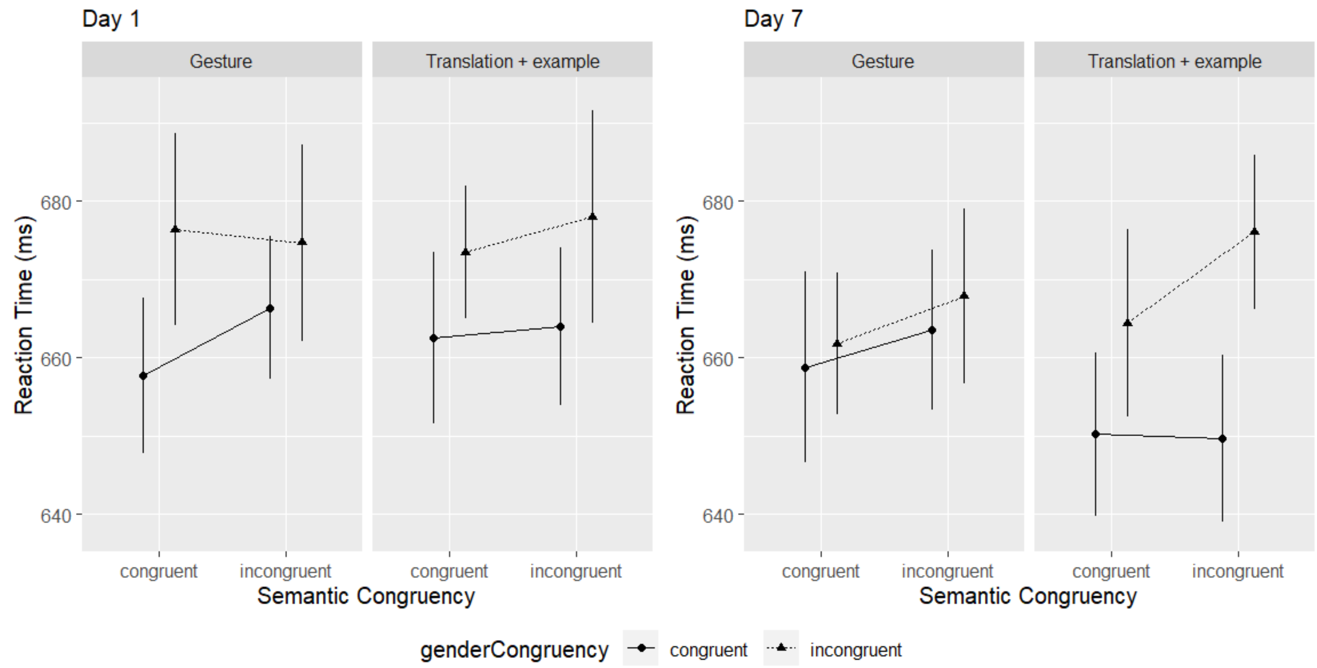


Figure 14. *Graph displaying the reaction times for the Action-speech integration assessment across the factors day, learning method, semantic congruency and gender congruency. Within-subject error bars reflect the 95% confidence intervals.*

Discussion

Study 2 aimed to investigate if the disambiguation account can, in part, explain the gestural advantage in L2 word learning as well as assess if the effects of gesture observed in the implicit task in Study 1 can be replicated. Overall, Study 2 found no differences between the gesture condition and the translation plus example condition in the explicit translation task or the implicit action-speech integration task.

Consistent with the disambiguation hypothesis, the gestural advantage was absent: there were no differences between learning methods in the translation assessment. Translation accuracy decreased across both learning methods from the first to the second assessment. This lack of a difference across learning conditions, after providing the translation cue with additional contextual information equal to that offered by gestures, suggests that disambiguation must be involved in the gestural advantage in some form. If disambiguation was not partly behind the gestural advantage, then the matching of contextual information across the two cue types would not have impacted the effect in any way. This research provides support that gestures do indeed have disambiguating properties and that these extend to L2 words, not just L1 speech (Holle & Gunter, 2007).

In contrast to Study 1, a difference in recall for the translations of the L2 words across the two assessment sessions was observed in Study 2. There was a significant decline in memory of the L2 words between the translation assessment one day after learning and the next seven days after learning. Although a decline in memory is not surprising given there were no additional learning sessions, it is unusual given that participants recorded slightly better overall translation accuracy (indicating greater learning) in this study than the first, and yet no differences between days were seen in Study 1. Learning in Study 1 appears to have been more resistant to decline. One potential explanation for this difference in studies could be that participants in Study 1 engaged more with the learning sessions (e.g., saying the words out loud, performing the gestures). This involvement may have led to greater resistance to decline over the week. Although both studies were given the same instructions, both were conducted online due to the Covid-19 lockdowns and restrictions and so the learning session could not be controlled or monitored in this way. Although Study 2 collected end of study survey data from participants (with around 70% of participants stating that they did repeat the words out loud and 64% stating they enacted the gestures) no data of this nature was collected for Study 1 and so a comparison cannot be made. Nevertheless, the self-proclaimed involvement in learning from the participants in Study 2 appears quite high and so this difference in involvement across the studies may not be the case.

Also, in contrast to the first study, Study 2 found no semantic congruency effect for words learnt through gestures in the action-speech integration task. This difference between studies does not appear to be a result of lower levels of learning as performance was slightly higher during the translation task for Study 2 (*day 1 Mean percentage accuracy* = 59%) compared to Study 1 (*day 1 M* = 54%). This raises the question of why the same semantic congruency effect for the gesture learning condition was not found in Study 2, despite this learning method being the same as that used in Study 1.

Study 2 did have a far greater number of outliers identified in the action-speech integration task (15 compared to only three in Study 1) which could indicate that the participants in this study paid less attention to the instructions and thus completed the sessions to a poor standard. However, these participants were removed from analysis and so would not have impacted the results. Additionally, the participants whose data remained do appear to have understood the instructions and engaged in the task as significant gender congruency effects were found which would not have been evident if participants had, for example, looked away from the screen. Note that gender congruency effects are not always present in gesture-speech integration tasks (Kandana Arachchige et al., 2022) but that their presence can be indicative of participants' attentiveness to the task.

There must be an alternative explanation as to why the same effects were not found in the action-speech integration task for the gesture learning condition in Study 2. It is not clear from these results if the implicit semantic congruency effects are simply fragile and potentially inconsistent, or if they even exist at all at this early stage of language learning. The semantic congruency effects in this task will be monitored in the upcoming studies and the results closely considered to determine if any further explanation can be provided.

Despite not finding effects in the implicit assessment measure, the equal learning displayed in the translation task provides insight that the manipulation and combination of various cues could improve learning. As the addition of example text added to translations boosted learning, additional cues added alongside gestures may have the same effect. Since the example text cue and gesture cues provide the same disambiguating information, the addition of L1 translations alongside gestures may provide an extra level of disambiguation that is beneficial to learning. This will be explored in the upcoming studies.

To conclude, equal learning across cue conditions was evident when translation cues are given accompanying text that provided the same contextual information as the gesture cues. This indicates that the disambiguation account must be in some way involved in the gestural advantage found in L2 vocabulary learning. Such findings stimulate ideas for the potential combination of cues to boost disambiguation further. The effect of gesture on the action-speech integration task observed in Study 1 did not replicate in Study 2. This task will be scrutinised in the upcoming studies in an effort to determine an explanation for the variability of effects.

STUDY 3 – INVESTIGATION INTO THE LEARNING EFFECTS OF L1 TRANSLATION INCLUSION ON GESTURE CUES

Introduction

The results of Study 2 raise the question as to whether the effectiveness of gestures could be further enhanced through additional cues, in the same way that the use of L1 translation as a cue of L2 learning was improved through example text. As highlighted previously in the 'Summary of open questions and research aims' (p.41) section of the general introduction, the past literature largely presented gestures alongside L1 translations. In Study 1, it was established that the accompanying L1 translation was not required alongside the gesture cue during learning for the gestural advantage to emerge, and gave several practical explanations as to why the presentation of gestures alone may be beneficial in a learning environment. However, to the best of my knowledge, the use of gestures alone has not yet been compared to using gestures alongside L1 translations. In the Introduction to Study 1 (p.45), I alluded that gestures alone may have greater learning benefits due to the application of the RHM, but in the light of the results from Study 2, this is now to be questioned.

It is fair to assume that the addition of an L1 translation to a gesture cue would add an extra level of disambiguation. If there was any uncertainty in the learner about what the gesture was depicting, this would be resolved with the addition of L1 translation text. But the question remains, would this addition enhance or hinder the effectiveness of gestures in L2 word learning? There are several explanations that can be given to argue for either possible outcome.

Firstly, the inclusion of L1 text accompanying the gesture could be advantageous for learning by simply providing an additional cue to the L2 word. The more cues available during learning, the richer the encoding will be, thus providing greater recall. A similar explanation for why the addition of the L1 translation with gesture cues would benefit learning involves the application of the Dual Coding Theory (Paivio, 1990). Within the DCT, verbal and extralinguistic stimuli are encoded into separate stores. These stores are linked through referential connections. The DCT proposes that stimuli that are encoded into both the verbal and nonverbal store (two routes to encoding available) will promote greater learning

(Clark & Paivio, 1991). Including an L1 translation will provide a verbal cue as well as the extralinguistic gesture cue. It has also been proposed that gestures can provide a motor sensory modality route to the nonverbal store (Huang et al., 2019; Paivio, 1978). Application of this proposal would suggest that a gesture plus translation learning method would allow for three encoding routes to develop: verbal, visual and motor. Application of the bilingual DCT would also suggest that the addition of the L1 translation would benefit learning as a cue would then be encoded into each system (L2 system, L1 system and the shared image system, Paivio & Desrochers, 1980).

Another reasoning as to why additional disambiguating information may benefit gestures comes from the results of the previous study. Study 2 revealed that when the L1 translation was accompanied with example text that gave equivalent contextual information to the gesture cue, the gestural advantage disappeared. Based on this, and applying the privileged access account, one could argue that the reason for equivalent learning, as opposed to greater learning for the translation plus example condition, was because the gesture-based learning condition still contained gestures which has specialised access to motor traces. This therefore suggests that a learning method that includes both gestures and the L1 translation would have an even greater learning effect than that of sole gestures, as the learning method has privileged access in addition to advanced disambiguation.

However, this prediction conflicts with the initial application of the RHM to this research. The reasoning behind investigating the use of gestures alone during Study 1 was based on the RHM's theory of lexical and conceptual links from the L2. Inclusion of the L1 translation would appear to facilitate the longer lexical links between L2 and L1, compared to the exclusion of L1 information during learning and only using cues that provide conceptual information. Such a learning environment would promote the development of direct conceptual links from the L2. Based on this model, if gestures are accompanied by L1 translations, learning would be less advanced, with only strong lexical links developing, unlike when gestures are used on their own.

Another explanation that suggests gestures would not benefit from the addition of L1 translations comes from the levels of processing (LOP) framework for memory (Craik & Lockhart, 1972). This framework proposes that processing, i.e., the rapid analysis of stimuli, is comprised of stages that differ in depth of encoding. Levels of processing theory proposes that memory is greater for information that has been processed on a deeper, conceptual level compared to a shallower, perceptual level (Craik & Lockhart, 1972). Deeper processing relates to more meaningful analysis at a semantic level and therefore is relevant to L2 learning.

One concept outlined in the LOP framework with particular relevance to semantic processing is 'elaboration' (Lockhart & Craik, 1990). Elaboration can be used to explain why the way in which information is presented, or the tasks included during learning, can increase the enrichment of the information during encoding, thus improving memory. Elaborative processes can be aided by the active involvement of learners in the tasks and learning process, i.e., active rather than passive encoding (Craik & Tulving, 1975). These elaboration effects have also been found in implicit memory tasks (Nicolas et al., 1996).

This concept can be applied to the proposed potential advantage of gesture cues on their own as opposed to with accompanying L1 translations. Although gestures can provide disambiguating contextual information, it can be argued that, when presented on their own, they still require participants to make some inferences about their meaning. For example, in the gesture for 'pour', the participant must identify that the target information is not cup, juice or liquid but the verb 'to pour'. Without accompanying text, gesture learning cues require greater cognitive effort to make these interpretations and pick out the meaning from the gesture. This semantic elaboration may lead to improvements in memory for the L2 words being taught. The application of this theory to L2 vocabulary learning has been specifically investigated, with the active process of retrieval during learning leading to greater recall of L2 words (Barcroft, 2007).

Aims for Study 3

Overall, there are a number of possible explanations for why the addition of L1 translations to gesture cues could help or hinder the gestural advantage. This study will therefore be used to investigate these two alternative predictions by comparing a gesture only learning condition with a gesture plus translation condition. In addition, given the conflicting findings from the action-speech integration task in Study 1 and Study 2 (with no semantic congruency effects found for words learnt through the gesture learning condition in Study 2), this task will again be used in Study 3 to identify if the gestural advantage can be detected under such implicit speeded conditions. The main questions being investigated in this study are:

i. Will the inclusion of the L1 translation to gestures improve or weaken the usefulness of gestures as cues in L2 word learning?
ii. Does the effect of gesture on the implicit speeded task observed in Study 1 replicate?

Method

Participants

Participants were recruited through the Universities SONA system and were either offered 2.5 course credit and a £10 Amazon voucher or a £20 Amazon voucher. Data collection took place from October to December 2021.

In total, 56 participants enrolled in the study, however 4 did not start the learning stage, 3 had incomplete learning phase data, 1 completed the learning phase but technical issues meant that the data did not save, 4 completed only the learning stage and no assessment and 2 did not completed the final assessment session.

The remaining participant group of 42 that completed the study contained 27 females and the mean age of participants was 26.36 (*SD* = 9.29). Of these participants, 40 received course credit and the £10 Amazon voucher and 2 received the £20 Amazon voucher. For 30 of the participants, English was their first language. For the other 12 participants, all stated English as a fluent second language. None of the participants had any previous knowledge of any Chinese vocabulary prior to participation in the study. In a self-reported question on personal motivation to learn Chinese, 18 participants stated that they were personally motivated.

Design

The same within-participant design that was used in the previous two experiments was used for this third study. The translation task used a 2 x 2 factorial design, using the two factors Learning Method (gesture, gesture plus translation) and Day (1, 7). For the action-speech integration task, a 2 (learning method) x 2 (day) x 2 (semantic congruency) x 2 (gender congruency) design was used. The same phases in the learning session were used, the gesture familiarisation, Mandarin exposure phase and the learning phase. Within this session, participants learnt 12 of the Mandarin words through gestures and the other 12 words through gestures plus translations. This session was designed in the same way as the previous studies, with a total of 12 blocks and 96 mini-tests. The same explicit and implicit assessment sessions were also used 1 day and 7 days after learning took place.

Stimuli

The same 24 Mandarin words were used for this study, as well as the same gesture video clips for learning and action video clips for assessment. For this study, a gesture plus translation learning condition was created by presenting the gesture clip with the corresponding English translation text (shown in Figure 15).

Procedure

Participants signed up for the study via the University's SONA system. On the morning of their booked day, they would be emailed the link to the first session. Participants were randomly allocated to either gesture set A or B which determined which set of 12 Mandarin words were taught through each learning method.

Learning session

After completing the initial survey with the consent form, participants began the learning session. Participants completed a Mandarin exposure phase and a gesture familiarisation phase (only for the gestures in the gesture only learning condition) before the learning phase. During the learning phase, participants learnt 12 of the Mandarin words through gestures and 12 through gestures plus translations (see Figure 15).

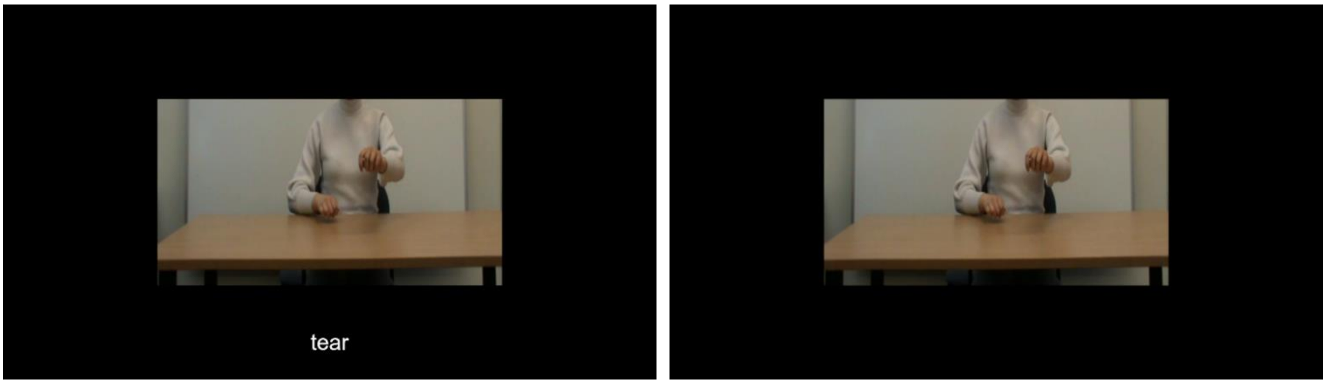


Figure 15. Freeze frames from the learning phase displaying the gesture plus translation learning cue (left) and the gesture learning cue (right).

Day 1 and Day 7 Assessment sessions

For the second session, participants recruited through SONA were sent their session link by 9am the day after learning was completed. The first assessment session (day 1) was completed, on average, 27.17 hours (*SD* = 7.35 hours, *min* = 16.54 hours, *max* = 55.65 hours) after learning took place. Four participants completed the first assessment session 1 day late (*M* = 44.83 hours after learning). These were still accepted as they were within the leeway time period.

The same process was used for the third session for participants that had completed session 2. Seven days after learning was completed, participants were sent their session link. This day 7 assessment was completed, on average, 173.28 hours (*SD* = 10.30 hours, *min* = 157.40 hours, *max* = 205.99 hours) after learning took place. Six participants completed this assessment session 1 day late (*M* = 193.84 hours after learning).

Transparency and Openness

This study was pre-registered online using AsPredicted. A copy can be found in the ResearchBox folder for this thesis (https://researchbox.org/2841) under the section 'Study 3'. Note that this study is named 'Study 2' in the pre-registration title. *Deviations from pre-registered protocol:* Only participants with a full data set were used in analysis as results were analysed using a 2x2x2x2 ANOVA using day as a factor and not analysed separately for each day.

Data Analysis

Mini-test performance

Mini-test data from the 48 participants that completed the full learning phase was gathered. Overall, for the 96 mini-tests, participants had an accuracy of over 90% (*M* = 89.23, *SD* = 6.20). Three participants were identified as having mini-test scores 2 *SD*s, or more, below the mean (<76.82) (see Appendix D, Table D3, p.168 for Study 3 mini-test data). Of these, one participant did not complete the day one assessment on time and so was stopped from completing the study and another only completed the day 1 assessment and not day 7. Therefore, of the 42 participants that completed the study, 41 remained for analysis (26 female*, M age* = 26.54, *SD* = 9.33).

Action-speech integration task outliers

Of the 41 participants, one participant only completed the practice for the day 1 action-speech integration task (session may have crashed or pressed escape by accident), this participant was removed from analysis.

Participants accuracy during the action-speech integration task was checked. For overall accuracy, one participant had less than 80% (72.40% overall, 50.52% for day 1, 94.27% for day 7). This participant's data was removed. All the participants in the data set now had over 80% accuracy overall and split by day. The lowest overall accuracy was 85.16% and when split by day, the lowest was 82.29%.

For the remaining 39 participants, outliers were then filtered out, with all trials less than 200ms being removed as pre-emptive trials, all reaction times over 2000ms being removed as 'timed-out' trials and all incorrect trials also being removed. Finally, individual participant outliers were removed (trials +/- 2.5 *SD*s from participants mean reaction times).

After these rejections, a total of four participants had less than 80% of trials remaining (two participants had too few trials on both days and two participants had too few remaining for day 7). These participants were removed from the whole data set, leaving 35 participants for analysis.

Results

Translation

For the first assessment day, in total the 41 participants correctly translated just over 40% of the Mandarin words (*Mean percentage accuracy* = 42.28%, *SD* = 24.57%). There were more words were correctly translated from the gesture plus translation learning condition (*M* = 46.75%, *SD* = 26.41%) than from the gesture learning condition (*M* = 37.80%, *SD* = 27.20%).

For the day 7 assessment, overall participants correctly translated just under 40% of words (*M* = 38.01%, *SD* = 23.83%). Again, more words were correctly translated in the gesture plus translation condition (*M* = 40.85%, *SD* = 24.57%) than in the gesture condition (*M* = 35.16%, *SD* = 25.58%).

A 2-way ANOVA was conducted using the factors day and learning method (see Figure 16). There was a significant effect of learning method, $F(1,40) = 8.32$, $p = .006$, $\eta^2_g = .020$, indicating that more words were correctly translated from the gestures plus translations condition (*Mean accuracy* = 0.438, *SE* = 0.0382) than words learnt through gestures (*M* = 0.365, *SE* = 0.0397). There was also a significant effect of day, $F(1,40) = 6.26$, $p = .017$, $\eta^2_g = .007$, indicating more correct responses on assessment day 1 (*M* = 0.423, *SE* = 0.0384) than on day 7 (*M* = 0.380, *SE* = 0.0372). There was no significant interaction between learning method and day, $F(1,40) = 1.23$, $p = .275$. Post-hoc tests indicated the effect of learning method was significant at the first assessment day (*M* = 8.94%, *SE* = 3.12%, *p* = .0042) but not the second (*M* = 5.69%, *SE* = 3.12%, *p* = .0680).

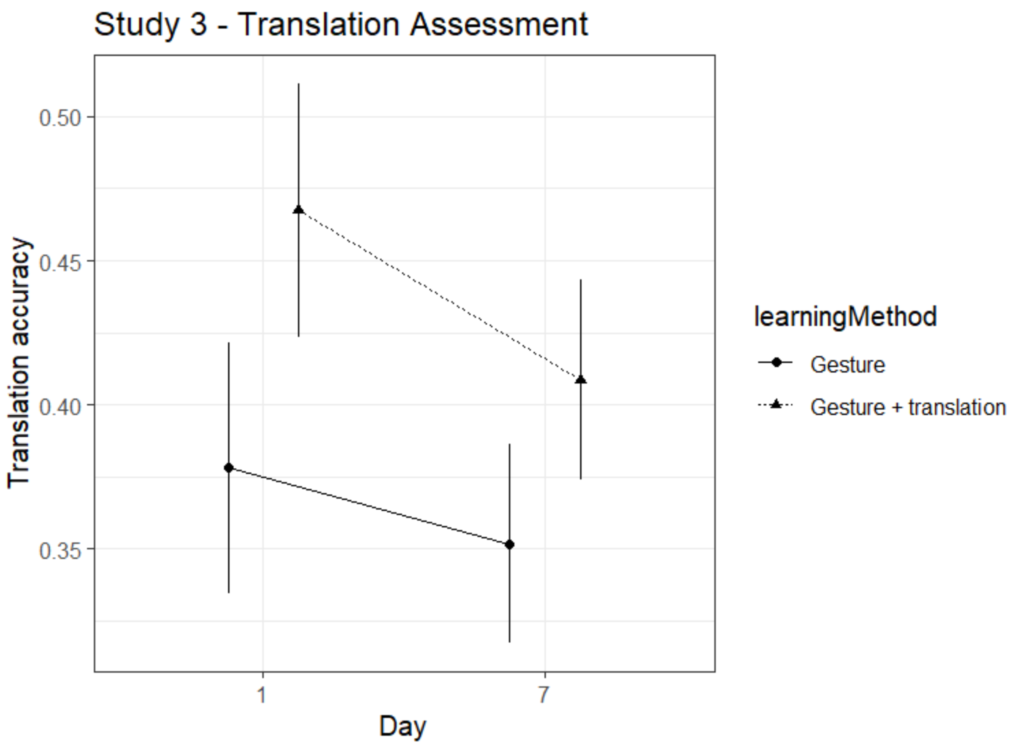


Figure 16. Graph displaying the mean translation scores across assessment day 1 and 7 for words learnt through gesture cues and gesture plus translation cues. The within-subject error bars reflect the variability of the data using 95% confidence intervals.

Action-speech integration task

The reaction time data was analysed using a 2 (learning method) x 2 (day) x 2 (semantic congruency) x 2 (gender congruency) ANOVA (see Figure 17). There was a significant main effect of gender congruency, $F(1,34) = 8.03$, $p = .008$, $\eta^2_p = .19$, indicating longer reaction times for gender incongruent (*M* = 722ms, *SE* = 26.7ms) than congruent trials (*M* = 710ms, *SE* = 28.1ms). A significant two-way interaction was observed between learning method and semantic congruency $F(1,34) = 8.58$, $p = .006$, $\eta^2_p = .20$, indicating that the size and direction of the semantic congruency effect was affected by learning method. For words learned via gestures plus translations, semantically incongruent trials produced longer reaction times (*M* = 719, *SE* = 27.9) than semantically congruent trials (*M* = 710, *SE* = 27.3). A reversed pattern was observed for words learned via gestures, where shorter reaction times were observed for semantically incongruent (*M* = 713, *SE* = 27.5) as compared to congruent trials (*M* = 722, *SE* = 27.5). All other main effects or interactions were not significant.

Although a three-way interaction of day × learning method × semantic congruency was not observed, it was additionally investigated whether there was an interaction between learning method and semantic congruency within each assessment day, to see whether it is robust within each assessment occasion. For the day 1 assessment, there was no significant interaction between semantic congruency and learning method, $F(1,34) = 0.469$, $p = .4980$. For the day 7 assessment, there was a significant interaction between semantic congruency and learning method, $F(1,34) = 8.705$, $p = .0057$.

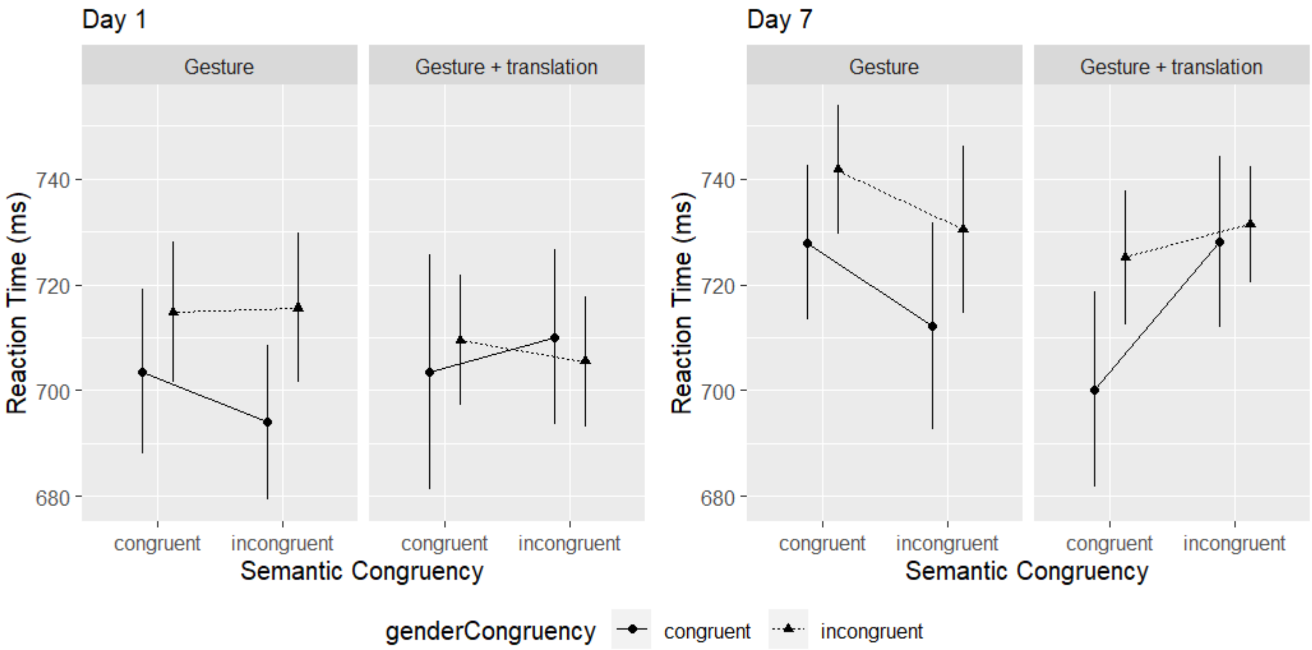


Figure 17. Graph displaying the reaction times for the Action-speech integration assessment across the factors day, learning method, semantic congruency and gender congruency. Within-subject error bars reflect the 95% confidence intervals.

Discussion

Study 3 aimed to identify if the inclusion of L1 translations alongside gesture cues would have beneficial or adverse effects on L2 word learning. Greater translation accuracy was found when gestures were accompanied by L1 text. Findings from the implicit action-speech integration task continue to conflict with Study 1.

This study found that more correct translations were made for words learnt through gestures plus L1 translations than from gesture alone. This indicates that although the gesture advantage does persist when gestures are the only cue in learning (as found in Study 1), gestures are an even more effective learning cue when accompanied with the L1 translation. These results are conflicting with my initial interpretation of what the RHM would predict. Based on the lexical and conceptual links from the L2, it was believed that limited exposure to the L1 translation during gestural learning would elicit development of conceptual links and thus result in greater learning via this method. However, as discussed previously in the Introduction to Study 3 (p.78), there are two possible explanations for the advantage of the gesture plus translation learning condition, and they are not necessarily mutually exclusive.

Firstly, the addition of the L1 text to the gesture may be providing an extra level of disambiguation. Study 2 suggested that additional cues can add disambiguating information, and that was applied in Study 3 with the addition of L1 translations removing any doubt in the learners' mind as to what the gesture was representing. However, from these results it is not clear whether the advantage found is solely due to the greater level of disambiguation in the gesture plus translation condition or whether the number of cues is also impacting results. A second explanation, based on the DCT (Paivio, 1990), is that the gesture plus translation learning condition activates multiple encoding routes, with both a verbal and an extralinguistic cue being provided, and thus information is inputted into both stores. This is compared to the gesture only learning condition which would only be encoded into the nonverbal store.

Showing inconsistency with Study 1, in the action-speech integration task this study found reverse congruency effects for the gesture only condition, and found these effects on day 7 instead of day 1. Unexpectedly, semantically incongruent stimuli had shorter reaction times than semantically congruent stimuli for the day 7 assessment. As with Study 2, a significant gender congruency effect was found, indicating that participants were engaged in the task and did not struggle with completing it correctly. These results question the reliability of the action-speech integration task to measure semantic congruency in early learners.

However, learning in the explicit task was notably lower for this study compared to the first two studies. Therefore, if the semantic congruency effects being measured in this task are fragile, as proposed following Study 2, then it is not surprising for these to be absent when learning is reduced as it is in Study 3. Moreover, responses were considerably longer in this study overall compared to the first two, with reaction times around 40ms slower. This may reflect the lower levels of learning demonstrated in the translation task. This implicit assessment will continue to be reviewed in the following studies.

Overall, Study 3 has highlighted the role of disambiguation in the gestural advantage. However, it has also revealed the impact that the number of cues can also have on learning. To truly understand the gestural advantage and be able to investigate the other possible mechanisms involved, learning methods with equal cues and disambiguation should be compared. Now that the effect of disambiguation has been established, this will allow for the privileged access account to be investigated in isolation.

STUDY 4 – THE ROLE OF THE PRIVILEGED ACCESS ACCOUNT IN THE GESTURAL ADVANTAGE

Introduction

Up to this point, the studies conducted have not been able to individually test privileged access due to the learning conditions used varying in other factors that have the potential to impact learning. Privileged access is the theory that proposes gestures allow specialist access to action representations and the motor system that are not provided through other learning cues. In Study 1 (gesture vs translation cues), despite having an equal number of cues across learning conditions, both privileged access and disambiguation could explain the results. In Study 2 (gesture vs translation plus example cues), learning conditions contained an unequal number of cues. Additionally, the level of disambiguating information was still potentially uneven across the two learning methods. Finally, Study 3 (gesture vs gesture plus translation cues) compared two gesture-based learning conditions and was therefore not investigating the privileged access mechanism, but did find that additional cues and extra disambiguating information during learning have a beneficial impact on L2 learning.

Thus, in order to finally be able to directly test the privileged access account for the gestural advantage in L2 word learning, a comparison is needed where a gesture and a non-gesture learning condition are compared, with all other third variables controlled for. In particular, both learning conditions should contain equal amounts of disambiguation and an equal number of cues.

Study 4 therefore compared gesture and non-gesture learning conditions with an equal number of cues and disambiguating information. A gesture plus translation learning condition was compared with a translation plus example condition, allowing the privileged access mechanism to be tested. Not only do these learning methods contain an equal number of cues, but they have also been found to lead to the best learning outcomes in gesture-based or translation-based learning conditions in the previous studies. Each learning condition contains equal disambiguating information, as both include an L1 translation and additionally a second cue that provides an example of the verb use (either in form of gesture or in the form of example text).

Despite not finding consistent semantic congruency effects in the action-speech integration task across the previous studies, the same explicit and implicit assessment measures were still used in this fourth study. As the action-speech integration task did produce the expected effects for the gesture learning condition in Study 1, the reliability of these results will continue to be tested in Study 4. This is particularly relevant given that this fourth study was the first to be conducted in the laboratory, as opposed to online following the Covid lockdowns. This may provide the action-speech integration task with further reliability as a measure of semantic learning.

Aims for Study 4

To summarise, the past studies have been unable to directly test the privileged access account for the gestural advantage in L2 word learning. This study implemented two learning conditions that were balanced in cues and disambiguating information. The key difference between the conditions was therefore the availability of privileged access provided by the gesture condition. The key questions being addressed in this study are:

i. Is the gestural advantage still evident when there is equal disambiguation and cues across learning conditions, as predicted by the privileged access account?
ii. Does the effect of gesture on the implicit speeded task observed in Study 1 replicate?

Method

Participants

Participants were recruited via the University of Hull's SONA Participant Sign Up website and were given 2.5 research course credits and a £10 Amazon voucher for completion of the study (42 participants). A further 4 participants were students that were known to the principal researcher and were recruited directly. These participants were awarded a £10 Amazon voucher on completion. Data collection took place from October 2022 to February 2023.

In total, 46 participants completed enrolment for the study, however 1 had no assessment session. The remaining participant group of 45 that completed the study contained 34 females and 11 males. The mean age of participants was 22.29 (SD = 6.99). For 35 participants, English was their first language. For the other 10 participants, all stated English as a fluent second language. None of the participants had any previous knowledge of any

Chinese vocabulary prior to participation in the study. In a self-reported question on personal motivation to learn Chinese, 15 participants stated that they were personally motivated.

Design

This study used the same design as the previous studies, the translation task used a 2 x 2 factorial design, using the two within-subject factors Learning Method (gesture plus translation, translation plus example) and Day (1, 7). For the action-speech integration task, a 2 (learning method) x 2 (day) x 2 (semantic congruency) x 2 (gender congruency) design was used.

However, unlike the previous studies that were conducted, this study was completed in person. Study sessions were completed in one of the Psychology computer laboratories on the University of Hull campus. The laboratory was a small room comprised of two desks with desktop computers. Due to participants being encouraged to repeat the Mandarin words out loud and perform the gestures, participant sessions were competed one at a time.

Stimuli

The same 24 Mandarin words were used for this study, as well as the same gesture video clips for learning and action video clips for assessment. In this study, participants learnt 12 of the Mandarin words through gestures plus translations and the other 12 words through translations and examples.

A short end of study survey was also used in this study. This survey was updated from the survey used in Study 2 to gain greater clarity and detail on certain areas covered in the questions. Answers for some questions (e.g., looking away from the screen during the action-speech integration task) were reported on a quantitative scale so that exclusions for data analysis could be made based on responses.

Procedure

Participants signed up to the study on the SONA system. They booked on to all three sessions at the same time, choosing which dates and timeslots worked for them.

Learning session

For their first session, participants would meet the researcher in the laboratory. After introductions, the researcher would run through the procedure of the session and answer any questions. After completing the initial survey with the consent form, participants would begin the learning session. Participants sat at the desk with over the ear headphones on. Participants completed a Mandarin exposure phase and a gesture familiarisation phase before moving on to the learning phase. During the learning phase, participants learnt 12 of the Mandarin words through gestures plus translations and 12 through translations plus examples (see Figure 18).

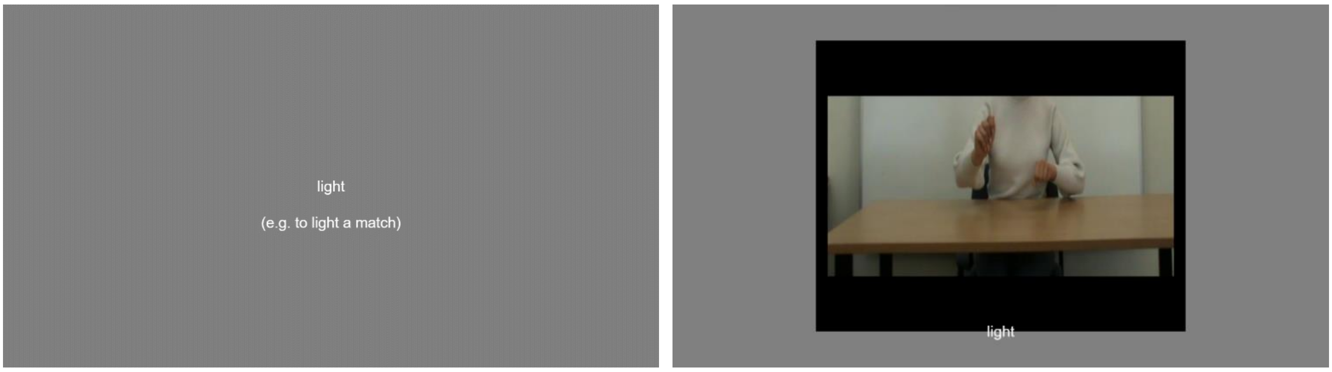


Figure 18. Freeze frames from the learning phase displaying the translation plus example learning cue (left) and the gesture plus translation learning cue (right).

Day 1 and Day 7 Assessment sessions

After the learning phase, participants would return for their scheduled timeslots 1 day and 7 days after learning for their assessment sessions. The first assessment session (day 1) was completed, on average, 23.12 hours (*SD* = 3.02 hours, *min* = 16.37 hours, *max* = 31.79 hours) after learning took place.

Seven days after learning was completed, participants would return for their final timeslot. This day 7 assessment was completed, on average, 167.37 hours (*SD* = 5.21 hours, *min* = 160.33 hours, *max* = 194.02) after learning took place. As in Study 2, an end of study survey was used in this study. After the final assessment session, participants were automatically redirected to the short survey.

Transparency and Openness

This study was pre-registered online using AsPredicted. A copy can be found in the ResearchBox folder for this thesis (https://researchbox.org/2841) under the section ‘Study 4’. Note that this study is named ‘Study 5’ in the pre-registration title. *Deviations from pre-registered protocol:* Only participants with a full data set were used in analysis as results were analysed using a 2x2x2x2 ANOVA using day as a factor and not analysed separately for each day.

Data Analysis

Mini-test performance

Mini-test data from the 46 participants that completed the full learning phase was gathered. Overall, for the 96 mini-tests, participants had an accuracy of over 95% (*M* = 91.41, *SD* = 4.21). Two participants were identified as having mini-test scores 2 *SD*s, or more, below the mean (<82.995) (see Appendix D, Table D4, p.168, for Study 4 mini-test data) and were removed. Therefore, of the 45 participants that completed the study, 43 remained for analysis (33 female, *M age* = 22.42, *SD* = 7.13).

Action-speech integration task outliers

Of the 43 participants, one participant did not have data saved for the day 1 action-speech integration task and was therefore removed from the analysis.

The accuracy, mean reaction times and standard deviations of the 42 participants remaining for analysis were checked. All participants had an accuracy of over 80%, both overall and when split by day, in the action-speech integration task. The lowest overall accuracy was 85.16% and when split by day, the lowest was 80.73%.

Outliers were then filtered out, with all trials less than 200ms being removed as pre-emptive trials, all reaction times over 2000ms being removed as 'timed-out' trials and all incorrect trials also being removed. Finally, individual participant outliers were removed (trials +/- 2.5 *SD*s from participants mean reaction times). After these rejections, a total of two participants had less than 80% of trials remaining, both with too few trials remaining for day 7. These participants were removed from the whole data set, leaving 40 participants for analysis.

The end of study survey data was reviewed to identify participants that did not follow the study instructions of keeping their eyes on the screen during the action-speech integration task. In total, a further eight participants were removed for looking away from the screen for over 10 of the videos (three looked away on both days, two looked away on day 1 and three looked away on day 7). Therefore, 32 participants remained for analysis.

Results

Translation

For the first assessment day, in total the 43 participants correctly translated just over 34% of the Mandarin words (*Mean percentage accuracy* = 34.11%, *SD* = 19.31%). There were more words were correctly translated from the gesture plus translation learning condition (*M* = 37.60%, *SD* = 21.55%) than from the translation plus example learning condition (*M* = 30.62%, *SD* = 22.18%).

For the day 7 assessment, overall participants correctly translated just over 32% of words (*M* = 32.85%, *SD* = 19.12%). Again, more words were correctly translated in the gesture plus translation condition (*M* = 35.85%, *SD* = 20.13%) than in the translation plus example condition (*M* = 29.84%, *SD* = 22.87%).

A 2-way ANOVA was conducted using the factors day and learning method (see Figure 19). There was a significant effect of learning method, $F(1,42) = 5.27$, $p = .027$, $\eta^2_g = .022$, indicating that more words were correctly translated from the gesture plus translation condition (*Mean accuracy* = 0.367, *SE* = 0.0294) than from the translation plus example condition (*M* = 0.302, *SE* = 0.0329). There was no significant effect of day, $F(1,42) = 0.46$, $p = .500$, and no interaction between learning method and day, $F(1,42) = 0.16$, $p = .691$. Post-hoc tests indicated the effect of learning method was significant both at the first (*M* = 6.98%, *SE* = 2.93%, $p = .0175$) as well as at the second assessment day (*M* = 6.01%, *SE* = 2.93%, $p = .0407$).

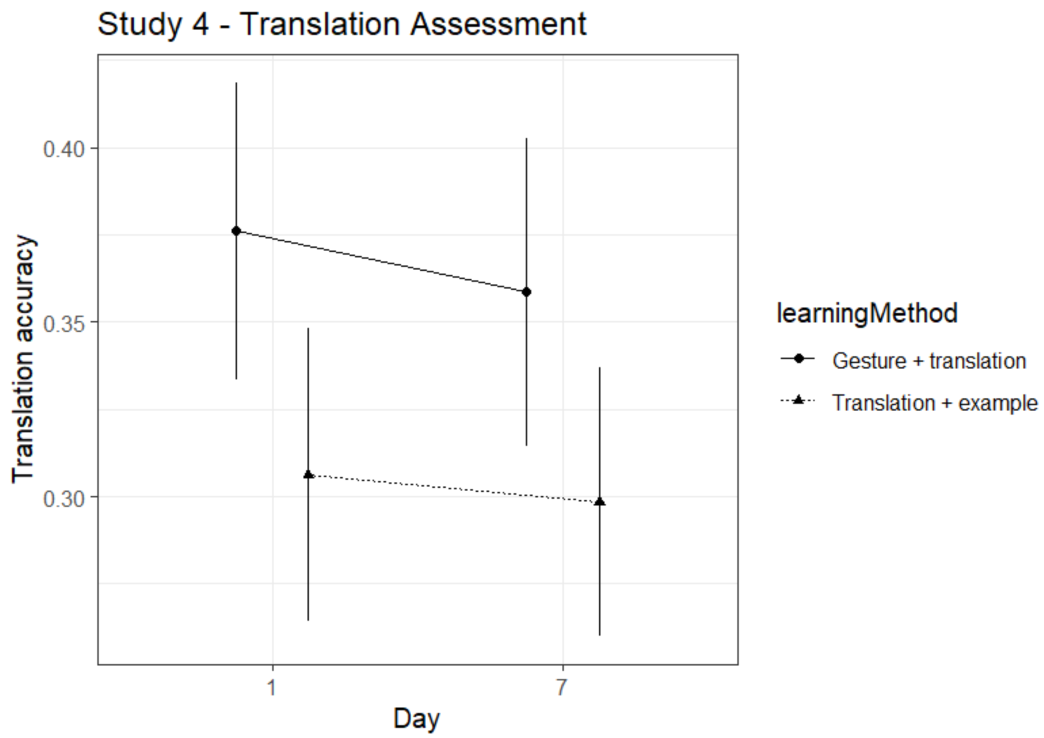


Figure 19. *Graph displaying the mean translation scores across assessment day 1 and 7 for words learnt through gesture plus translation cues and translation plus example cues. The within-subject error bars reflect the variability of the data using 95% confidence intervals.*

Action-speech integration task

The reaction time data was analysed using a 2 (learning method) x 2 (day) x 2 (semantic congruency) x 2 (gender congruency) ANOVA (see Figure 20). There was a significant main effect of gender congruency, $F(1,31) = 23.78$, $p < .001$, $\eta^2_p = .43$, indicating longer reaction times for gender incongruent (*M* = 625ms, *SE* = 20.1ms) than congruent trials (*M* = 608ms, *SE* = 19.3ms).

There was also a significant interaction between gender congruency and day, $F(1,31) = 6.15$, $p = .019$, $\eta^2_p = .17$, indicating that the difference in reaction times between gender congruent and incongruent trials was more pronounced for day 1 (*M congruent* = 609, *SE* = 18.0; *M incongruent* = 632, *SE* = 19.7) than for day 7 (*M congruent* = 606, *SE* = 23.5; *M incongruent* = 617, *SE* = 23.1). All other main effects or interactions were not significant.

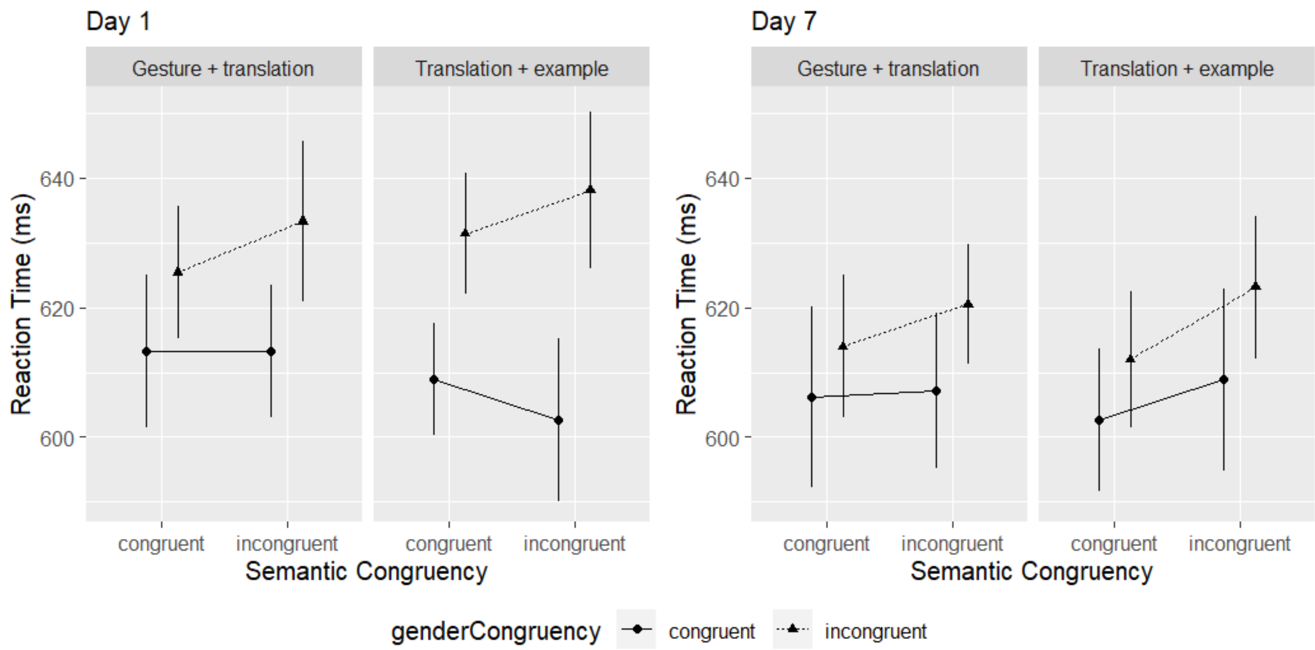


Figure 20. Graph displaying the reaction times for the Action-speech integration assessment across the factors day, learning method, semantic congruency and gender congruency. Within-subject error bars reflect the 95% confidence intervals.

Discussion

Study 4 aimed to investigate if the privileged access account is involved in the gestural advantage by testing the use of gesture cues when equal disambiguation and cues are used across the two learning conditions. Greater translation accuracy was recorded across both assessment days for words learnt through the gesture-based learning method. Again, no semantic congruency effects were found in the implicit action-speech integration task.

Study 4 revealed that when learning methods are controlled for by the number of cues provided and the level of disambiguating information, the gestural advantage still persists in L2 to L1 translations. This difference between a gesture-based and a non-gesture-based learning method persisted over time, with effects still evident 7 days after learning. This indicates that privileged access is involved in the gestural advantage in L2 vocabulary learning.

The gesture plus translation learning condition used in Study 4 provided a high level of disambiguation. In addition to this, the privileged access account proposes it also provides specialist access to motor traces and action representations via the gesture cue. Due to the close integration of gestures and speech, when the L2 words are encoded alongside iconic gestures, they are automatically integrated together. As a result, motor traces from performing, viewing someone performing, or imagining oneself performing gestures become associated with the accompanying L2 word. These motor traces are independent from other visual modality pathways that may become active with other extralinguistic cues such as pictures (Engelkamp, 1986). Gestures have found to still be advantageous over drawing the outline of pictures with one's finger (Mayer et al., 2015), demonstrating how these motor traces are unique to iconic gestures and not to other physical movements. During retrieval, when an L2 word is presented, these motor traces that have become associated with the words are activated, aiding the retrieval of conceptual information.

Although the previous literature highlighted the importance of enactment of gestures for the development of motor traces (Macedonia & Mueller, 2016), privileged access does not assume that this enactment is necessary. Enactment may be beneficial in further strengthening these traces, but research suggests that motor traces can still become activated through viewing gestures (Huang et al., 2019; Tellier, 2008), with greater learning over non-gesture cues still being recorded. Additionally, imagining oneself performing gestures may also activate similar motor traces (as found with actions, Nilsson et al., 2000; Nyberg et al., 2001). Overall, the privileged access account proposes that viewing and performing gestures accompanying L2 words activate motor traces that then become linked to the L2 words, thus aiding semantic retrieval.

The translation scores in Study 4 are low compared to those found in the previous studies. For the gesture plus translation condition, there was just under a 10% decrease in the day one translation scores found in the current study compared to those in Study 3. The only difference between the two studies, other than the other learning condition used, was that this study was conducted in-person as opposed to online. It may be that the laboratory setting, particularly for participants that had not experienced such an environment before due to the covid pandemic, created discomfort for them. The laboratory setting may have been seen as a more intense learning environment and as a result negatively impacted the scores (Véliz-Campos et al., 2023).

As with the previous studies, there was again no semantic congruency effects in the implicit action-speech integration task. Consistent with the previous studies, there was only an effect of gender congruency. Overall, the reaction times recorded were far quicker in this study than they were in any of the previous studies. All studies were run off of the Pavlovia server and so a technical difference between studies cannot explain the overall rection time differences. As this study was in person, participants may have felt more pressure due to the presence of the experimenter to respond quickly. These fast reaction times may have also been a sign that participants were following the task instructions more closely.

The lack of semantic congruency effects in this study calls into question why such inconsistent results within the action-speech integration task have been found throughout the series of studies presented so far. It has previously been suggested that the semantic effects on such a task may be fragile in early L2 learners. The use of multiple cues and learning conditions may create an environment too complex for deep semantic integration to develop. The next study addressed this issue in order to test if semantic congruency effects were detectable in the action-speech integration task for a final time when learning a smaller set of words through gestures only.

STUDY 5 – GESTURE REPLICATION: A FINAL TEST OF THE ACTION-SPEECH INTEGRATION TASK

Introduction

Over the past four studies inconsistent results have been found in the implicit action-speech integration task used to measures semantic integration of L2 words. After obtaining promising results in pilot studies, in which semantic congruency effects were found with L1 words in native and non-native speakers, the task appeared to be an effective measure. This seemed to be confirmed by the semantic congruency effects measured for words learnt through gestures in Study 1. However, such an effect for the gesture condition could not be replicated in Studies 2-4. In fact, in Study 3, a semantic congruency effect in the opposite direction was found in the gesture only learning condition, with semantically congruent trials having longer reaction times than semantically incongruent trials.

There are two possible explanations as to why the action-speech integration task has not produced a stable effect. The first is that semantic effects under such conditions simply do not exist this early on in L2 vocabulary learning. The direct semantic links from L2 can develop eventually and are evident in proficient L2 speakers in the action-speech integration task but may not have developed after only one learning session (with low proficiency learners only showing limited access to direct conceptual information, Dufour & Kroll, 1995). Despite their advantageous properties, gesture cues may not be able to fast-track semantic learning of newly learnt words enough to display such effects in the given task. However, this explanation cannot account for the semantic congruency effect from Study 1.

Another possible explanation is that the inconsistent results may be a consequence of the task or study design. The semantic congruency effects in these early stages of L2 acquisition could still exist but may be fragile and weak compared to those found in native and proficient speakers. The task contains both speeded and implicit elements which may make the task too demanding to allow the detection of any weak semantic connections in early L2 learners. Additionally, the study designs throughout the previous studies were rather complex, with two within-subject learning conditions, often using multiple cues, and two separate sets of L2 words used. Participants may have faced interference from the other learning conditions and their cues. Furthermore, the studies involved learning a total of 24 L2 words. This may be too many words for a single learning session to enable deep semantic learning to take place. A more simplified study design that has the potential to allow for greater learning may be needed to demonstrate semantic congruency effects.

The plan for this study was therefore to test a simplified study design. If semantic congruency effects are still not evident, attention can then be moved on to questioning the task itself. Study 5 therefore acts as a last attempt to test the action-speech integration task. A simplified learning session was used, with the intention of creating less confusion and greater learning. Only one learning condition was used containing gestures as the sole cue to the L2 word meanings. By doing so, only one cue modality was used for all words. Furthermore, this was the learning condition that originally resulted in semantic congruency effects in Study 1. Also, by including fewer L2 words, it provided greater opportunity for the word to be learnt on a deeper, sematic level.

Aims for Study 5

To recap, this study aims to investigate one of the possible explanations for why inconsistent results have been found in the action-speech integration task. This study will provide a final attempt to understand the significant result found in Study 1. The key question being explored is:

i. Is a more simplified study design, that allows for greater learning, needed for the effective measure of semantic congruency effects in newly learnt words through the action-speech integration task?

Method

Participants

Participants were recruited through Prolific and were paid as close to £8/hr as possible (payment was set to £8.12/hr for session 1 and £8.02/hr for session 2 and 3). These payments were set to the predicted average time to complete and may have fluctuated slightly based on the participant sample median completion time (session 1 – predicted 30 minutes, *Mdn* = 31.20, session 2 – predicted 13 minutes, *Mdn* = 14.58, session 3 – predicted 25 minutes, *Mdn* = 18.05). The study was only visible to participants within the UK and with a minimum of 10 previous submissions. The study not visible to those fluent in Mandarin, Cantonese or Chinese. Data collection took place from July to September 2022.

In total, 57 participants enrolled in the study, however one participant timed out (was inactive for too long) during the learning phase (no mini-tests completed), two participants returned to Prolific with incomplete learning phases, one participant completed only the learning phase as they were stopped from continuing due to only scoring just over 50% in the mini-tests, two completed only the learning stage and no assessment and one did not complete the final assessment session.

The remaining participant group of 50 that completed the study contained 28 females and the mean age of participants was 36.42 (SD = 11.90). For 46 of the participants, English was their first language. For the other four participants, all stated English as a fluent second language. None of the participants had any previous knowledge of any Chinese vocabulary. In a self-reported question on personal motivation to learn Chinese, 30 participants stated that they were personally motivated.

Design

This online study used a similar design to the previous online studies, however this study only used one learning method. The same gesture only learning condition from Study 1 was used to teach participants Mandarin words in this study. The translation task used a single factorial design using the within-subject factor Day (1, 7). For the action-speech integration task, a 2 (day) x 2 (semantic congruency) x 2 (gender congruency) design was used.

Stimuli

The Mandarin words, video and audio stimuli used in this study remained the same as the previous studies. In an effort to ensure a good level of learning would take place, only 12 Mandarin words were taught to participants. To choose which 12 words to use, the translation accuracy and semantic congruency effects of the words and word pairs from the gesture learning condition for the first 3 studies were reviewed (see Appendix E, p.177, for example data). Word pairs that previously displayed the strongest semantic congruency effects and translation accuracy scores were chosen.

Procedure

Participants signed up via Prolific and began the session on the same day that they signed up.

Learning session

During the learning phase, participants complete the Mandarin exposure phase, the gesture familiarisation phase and then the learning session. During the learning session, participants learnt the 12 Mandarin words through the gesture learning method (see Figure 21) and completed a total of 48 mini-tests.

Figure 21. Freeze frame from the learning phase displaying the gesture learning cue.

Day 1 and Day 7 Assessment sessions

For the second session, participants were invited to a 'Session 2 Study' by 9am on the morning of their second day. When added, participants received an email notification. The first assessment session (day 1) was completed, on average, 21.20 hours (*SD* = 5.85 hours, *min* = 16.03 hours, *max* = 53.05 hours) after learning took place. One participant completed the first assessment session 1 day late (53 hours after learning). This was still accepted as it was within the leeway time period.

Seven days after learning was completed, participants that had completed the first and second session were added to a 'Session 3 Study'. This day 7 assessment was completed, on average, 164.97 hours (*SD* = 5.68 hours, *min* = 160.05 hours, *max* = 190.56) after learning took place. Two participants completed the day 7 session 1 day late (190 hours and 182 hours after learning). These were still accepted as they were within the leeway time period. After the action-speech integration task in session 3 (day 7), participants were automatically redirected to the end of study survey.

Transparency and Openness

This study was pre-registered online using AsPredicted. A copy can be found in the ResearchBox folder for this thesis (https://researchbox.org/2841) under the section 'Study 5'. Note that this study is named 'Study 4' in the pre-registration title. *Deviations from pre-registered protocol:* Only participants with a full data set were used in analysis as results were analysed using a 2x2x2 ANOVA using day as a factor and not analysed separately for each day. Additionally, any participant that stated in the end-of-survey that they looked away from the screen for more than 10 videos was excluded from analysis.

Data Analysis

Mini-test performance

A total of 54 participants completed the learning phase mini-tests. On average, participants scored over 95% in the mini-tests (*M* = 45.65, *SD* = 3.88). Two participants had mini-test scores 2 *SD*s, or more, below the mean (<37.90) (see Appendix D, Table D5, p.168, for Study 5 mini-test data). One of these had already been stopped from completing the study after extremely low mini-test scores. The other was removed from the data set leaving 49 participants for analysis.

Action-speech integration task outliers

The accuracy, mean reaction times and standard deviations of the 49 participants remaining for analysis were checked. All participants had an accuracy of over 80%, both overall and when split by day, in the action-speech integration task. The lowest overall accuracy was 84.38% and when split by day, the lowest was 80.21%.

Outliers were then filtered out, with all trials less than 200ms being removed as pre-emptive trials, all reaction times over 2000ms being removed as 'timed-out' trials and all incorrect trials also being removed. Finally, individual participant outliers were removed (trials +/- 2.5 *SD*s from participants mean reaction times).

After these rejections, a total of two participants had less than 80% of trials remaining, one had too few remaining on day 1 and the other on day 7. These participants were removed from the whole data set, leaving 47 participants remaining.

Next, the end of survey data was reviewed and participants that looked away for more than 10 videos during one assessment session were removed. In total, six participants were removed, leaving 41 participants for analysis.

Results

Translation

For the first assessment day, on average the 49 participants correctly translated just over 64% of the Mandarin words (*Mean percentage accuracy* = 64.63%, *SD* = 25.60%). For the day 7 assessment, on average the 49 participants correctly translated just over 54% of the Mandarin words (*M* = 54.93%, *SD* = 29.65%).

The data was combined across days. An ANOVA was conducted on the combined data set and a significant effect of day was found, $F(1,48) = 13.8$, $p < .001$, $\eta^2_g = .030$.

Action-speech integration task

The reaction time data was analysed using a 2 (day) x 2 (semantic congruency) x 2 (gender congruency) ANOVA (see Figure 22). There was a significant main effect of gender congruency, $F(1,40) = 6.94$, $p = .012$, $\eta^2_p = .15$, indicating longer reaction times for gender incongruent ($M = 735$ms, $SE = 22.2$ms) than congruent trials ($M = 722$ms, $SE = 21.4$ms).

There was no significant effect of semantic congruency, $F(1,40) = 0.02$, $p = .893$, or interaction between semantic congruency and day, $F(1,40) = 3.80$, $p = .058$.

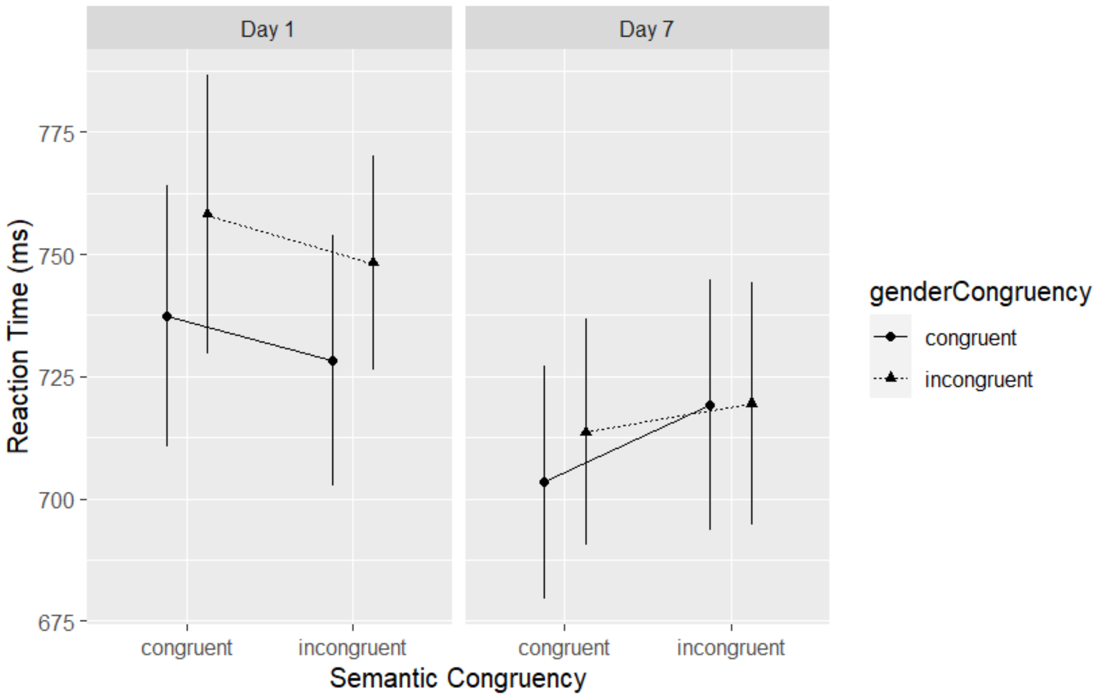


Figure 22. Graph displaying the reaction times for the Action-speech integration assessment across the factors day, semantic congruency and gender congruency. Within-subject error bars reflect the 95% confidence intervals.

Transparency and Openness

The pre-registration stated that "exploratory analysis of the relationship between the translation accuracy and the semantic congruency effect" would be conducted. However, as there was no significant semantic congruency effect found, this anlysis was not performed.

Discussion

This study failed to find semantic congruency effects in the implicit action-speech integration task for the gesture learning condition. This was despite the learning being simplified, with fewer L2 words being taught and only using one learning method. These changes were implemented in an effort to boost learning and thus provide the best chance for words to display semantic integration. These changes also aimed to eliminate any distractors or factors (such as multiple learning conditions) that made the learning session more complex or that could affect learning.

Learning did appear to be improved in the translation task, with this study producing the highest average percentage accuracy for day 1 across all the studies at roughly 64% (compared to approximately 59% in Study 2, and 58% in Study 1, for the gesture only condition). However, despite this, semantic congruency effects in the action-speech integration task were still not found. The only significant effect was again gender congruency, suggesting that the task was being attended to correctly.

This study confirms that, if any semantic integration has developed between the L2 words and concepts, it is not detectable in the action-speech integration task (see Appendix F, p.178, for results from this task for Studies 1-5). As discussed in the introduction to Study 5, two possible explanations remain. It could be that these semantic connections between L2 words and concepts do not exist at such an early stage of L2 learning. These connections will develop at some point, as suggested in previous literature using low proficiency L2 learners (Dufour & Kroll, 1995; Sunderman & Kroll, 2006), but perhaps not this early, despite the use of gesture learning cues. However, as highlighted previously, this explanation cannot explain the results of Study 1. It may be the case, though, that the semantic effects recorded in Study 1 were a statistical fluke.

The second explanation lies with the action-speech integration task itself and its potential limitations when used on very early L2 learners. The new connections that may be forming between L2 words and concepts could be fragile and not robust enough to display consistent semantic effects in a speeded implicit task (Kroll & Tokowicz, 2001). Research from a speeded cross-language categorisation task demonstrated that less-fluent bilinguals were significantly impacted (longer reaction times) by a mis-match in categorisation and target language (Dufour & Kroll, 1995). This indicates that although they were able to access conceptual information from the L2, it was limited. These findings suggest that even in low level proficiency bilinguals, L2 semantic connections may be present, but weak. A less challenging task may be more suited to investigate whether direct semantic connections can be detected in early L2 vocabulary acquisition.

STUDY 6 – TESTING THE GESTURAL ADVANTAGE UNDER SPEEDED, EXPLICIT CONDITIONS

Introduction

As previously highlighted, apart from a one-time effect that was not replicable, no evidence of learning in the implicit, speeded task has been found for the first assessment day. Additionally, no semantic congruency effects were found when a simpler learning environment was created that involved fewer L2 words and only one learning method.

Thus, attention then turned to the type of task being used to assess deeper semantic effects. It may be the case that after a single learning session, learning is not advanced enough to be evident in a speeded, implicit task. These two components of the action-speech integration task may create an assessment that is too demanding to allow any semantic learning effects, in these early stages, to be displayed.

Therefore, despite the previous studies having been conducted, questions remained as to whether the impact of the gesture advantage is limited to an explicit translation task or whether the benefit can extend beyond this controlled retrieval level in early L2 learners. To investigate this the first step is to consider the type of task being implemented. In Studies 1-5, the two assessment tasks used were:

- An un-speeded and explicit translation task
- A speeded and implicit action-speech integration task

This raises the possibility that a task that lies somewhere between the two, as a middle ground, could be an effective measure that may give further insight into the extent of the gesture advantage. Learning may be advanced enough to be evident in a speeded explicit task. Unlike the explicit un-speeded translation task used in our previous studies, a speeded explicit task would require quick, timed responses, whilst still using an explicit semantic judgement.

Relatedness Judgement task

A new outcome measure was therefore used for this final study, in addition to the translation task. A speeded, semantic relatedness judgement task was developed from similar tasks that have been used in the novel word learning literature (Dittinger et al., 2016, 2019).

In Dittinger et al.'s (2016) study, novel noun words were taught using picture-word associations. Learning was then assessed using a matching task and a semantic task, in which picture and audio stimuli of the novel words were presented simultaneously. In the matching task, the pictures were either the same as those presented with the words during learning or different. In the semantic task, the pictures either related to the novel word or were unrelated. For example, for the novel word taught as 'key', a related trial would present a picture of a lock. The related and unrelated images used during the assessment task had not been seen during the previous learning phase.

These tasks were adapted and applied to this study using video clips instead of pictures. Action videos were used during the task and played alongside either a matching or non-matching audio of a Mandarin word that had been taught. Despite depicting the same verb as those taught during learning, the action videos used in this task had not been viewed before, unlike the pictures used in Dittinger et al. (2016) matching task, thus giving this relatedness judgment task similar properties of that of their semantic task. The participants' task was to indicate as accurately and as quickly as possible whether the video and Mandarin word named were related or unrelated.

The most effective gesture and non-gesture learning methods from the previous studies were chosen for this study. The conditions from Study 4, gesture plus translation and translation plus example, were used due to the strong gestural advantage having been displayed with these conditions previously in the translation task. Additionally, the chosen learning conditions contain equal cues and disambiguation.

As well as implementing the new relatedness judgment task, this study also aimed to overcome some methodological issues that were not considered when these learning conditions were previously used. The main issue was that, in an effort to make the studies as consistent as possible with each other to aid cross-study comparisons, the gesture familiarisation phase was always included, despite not always being necessary. For Study 4, in which the gesture condition included L1 translations, the gesture familiarisation phase was still used before learning. This oversight led to additional exposure to the gestures and the L1 translations that may have contributed to improvements in learning. Therefore, no gesture familiarisation phase was used in this sixth study. Other, smaller changes were made from Study 4 that involved ironing out minor differences in the mini-tests between the gesture-based and the translation-based learning conditions. In Study 4, for the translation-based learning condition, only the L1 translation was used as the cue during the mini-tests and so this was changed to include the example text as well. Additionally, this text was made timeless so that it would remain on the screen until the participant chose to continue, to match the gesture condition.

Aims for Study 6

In summary, this study aimed to test a new outcome task that can measure semantic learning in a speeded, explicit environment. Due to the explicit nature of the task, it is more likely to detect semantic learning effects in early learners that were not consistently evident in an implicit task. The speeded component of the task enables the gesture advantage to be tested in conditions beyond a controlled translation task that may not be reflective of direct conceptual links from the L2. The following key question is explored:

i. Is the gestural advantage in L2 word learning only detectable with un-speeded explicit tasks, or does it generalise to speeded explicit tasks?

Method

Participants

Participants were recruited via the University of Hull's SONA Participant Sign Up website and were either given 1.5 research course credits and a £10 Amazon voucher for completion of the study (four participants) or signed up for £20 Amazon voucher (45 participants). Data collection took place between March and June 2023.

In total, 49 participants took part in the study. One participant did not complete the final assessment session. The remaining participant group of 48 that completed the study contained 32 females and 16 males. The mean age of participants was 27.75 (SD = 9.03). Of these participants, four received course credit and the voucher and 44 received just a voucher. For 40 participants, English was their first language. For the other eight participants, seven stated English as a fluent second language. One participant did not report English as a second language in the survey, but this was likely an error. As the participant was an undergraduate Psychology student at the University, it is fair to assume they are proficient in English. None of the participants had any previous knowledge of any Chinese vocabulary prior to participation in the study. In a self-reported personal motivation to learn Chinese, 23 participants stated that they were personally motivated.

Design

Two learning methods were used, with participants learning 12 of the Mandarin words using each method. The same learning methods from Study 4, gesture plus translation and translation plus example, were chosen as they have equal disambiguating information and cues. Additionally, a gesture plus translation advantage has already been demonstrated on the translation task using these conditions.

Learning Assessment Methods

Un-speeded learning assessment: translation task

In order to assess learning, two assessment tasks were used in this study. The first was the same explicit, un-speeded translation task used in the previous studies and described in Study 1. This translation task used a 2 x 2 factorial design, using the two within-subject factors Learning Method (gesture plus translation, translation plus example) and Day (1, 7).

Speeded learning assessment: relatedness judgement task

The second task was a speeded, explicit, semantic judgement task which used a 2 (learning method) x 2 (day) x 2 (relatedness) design. This relatedness judgement task was developed from similar tasks that have been used in the word learning literature (Dittinger et al., 2016; Dittinger et al., 2019). In the task, the action video clips are played alongside either a matching or non-matching Mandarin audio of a word that has been taught. The 24 Mandarin words were paired together in the same way as in the previously used action-speech integration task. There was no gender incongruency in this task, with all trials being presented with both the male and female recordings. The task contained a total of 96 trials. Participants' task was to indicate as accurately and as quickly as possible whether the action video and the Mandarin word named were related or unrelated. Accuracy scores and reaction times were recorded.

Stimuli

The same Mandarin words, video and audio stimuli used in the previous studies were used in this study. The new speeded relatedness judgement task took place after the translation task in the assessment sessions. The task included 96 trials, of which half used female actors. Additionally, half of the trials used congruent action videos and Mandarin speech, and half incongruent.

During this task, action videos began to play and after 200ms the Mandarin speech would play. Participants were required to press either F or J to indicate if the stimuli were congruent or incongruent (counterbalanced across participants). A red warning message appeared 2000ms after the onset of speech to prompt participants to respond quicker (see Figure 23). There was no time out implemented and participants had to respond before continuing.

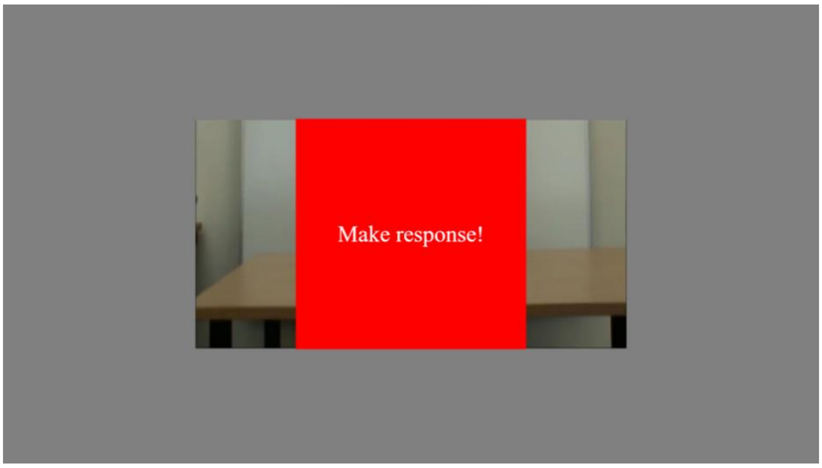


Figure 23. Freeze frame from the relatedness judgment task showing the warning message displayed after 2000ms if no response has been recorded.

Procedure

Participants signed up to the in-person study on the University of Hull's SONA system. The study was conducted on campus in the University Psychology laboratories.

Learning session

Participants first completed the learning phase individually in the laboratory. This comprised of the Mandarin exposure phase, and the learning phase. This session lasted about 45 minutes.

Day 1 and Day 7 Assessment sessions

The following day the participants came at their scheduled time to complete the first assessment session (translation task and the speeded relatedness judgement task). This session lasted about 15 minutes. The first assessment session (day 1) was completed, on average, 23.36 hours (*SD* = 2.77 hours, *min* = 16.58 hours, *max* = 30.73 hours) after learning took place.

Seven days after learning participants returned for their third session. This day 7 assessment was completed, on average, 168.41 hours (*SD* = 3.88 hours, *min* = 164.10 hours, *max* = 192.17) after learning took place. One participant completed the day 7 session 1 day late (192 hours). This was still accepted as it was within the leeway time period. After the translation and speeded relatedness judgement task, participants were automatically redirected to the end of study survey. This session lasted about 20 minutes in total.

Transparency and Openness

This study was pre-registered online using AsPredicted. A copy can be found in the ResearchBox folder for this thesis (https://researchbox.org/2841) under the section 'Study 6'. *Deviations from pre-registered protocol:* Only participants with a full data set were used in analysis as results were not analysed separately for each day. To view the correlation analyses of the relationship between the two learning measures, see Appendix G, p.180.

Data Analysis

<u>Mini-test performance</u>

Mini-test data from the 49 participants that completed the full learning phase was gathered. In the learning phase mini tests (out of 96), these participants had a mean score of 90.78 (*SD* = 6.73). Of these, four participants had mini-test scores 2 *SD*s, or more, below the sample mean (<77.31) (see Appendix D, Table D6, p.168, for Study 6 mini-test data) and so were excluded from analysis. Additionally, one participant did not complete the day 7 assessment and so was also excluded. Therefore, 44 participants (29 female, *M age* = 27.70, *SD* = 9.28) remained for analysis.

<u>Relatedness judgment task reaction time outliers</u>

Responses under 200ms were removed from the data set as these were classed as pre-emptive trials. Additionally, responses over 2000ms and reaction times above or below 2.5 standard deviations from the participants mean were also removed. Overall, 84.29% of the original trials remained in the whole data set. One participant only had 3 correct responses remaining for day 1 and so was removed during the ANOVA analysis of the data, leaving 43 participants in the relatedness judgement task data.

Results

Translation

For the first assessment day, on average the participants correctly translated just over 30% of the Mandarin words (*Mean percentage accuracy* = 31.53%, *SD* = 20.73%). On average, more words were correctly translated in the gesture plus translation condition (*M* = 32.77%, *SD* = 20.36%) than in the translation plus example condition (*M* = 30.30%, *SD* = 25.55%). For the day 7 assessment, participants correctly translated 35% of the Mandarin words (*M* = 35.04%, *SD* = 22.22%). On average, more words were correctly translated in the gesture plus translation condition (*M* = 38.07%, *SD* = 24.07%) than in the translation plus example condition (*M* = 32.01%, *SD* = 23.91%).

A 2-way ANOVA was conducted using the factors day and learning method (see Figure 24). There was a significant main effect of day, $F(1,43) = 5.19$, $p = .028$, $\eta^2_g = .006$, indicating more correct responses on day 7 (*Mean accuracy* = 0.350, *SE* = 0.0335) than on day 1 (*M* = 0.315, *SE* = 0.0313). There was no significant effect of learning method, $F(1,43) = 2.54$, $p = .118$, and no significant interaction between learning method and day, $F(1,43) = 2.51$, $p = .121$.

Post-hoc tests indicated the effect of learning method was not significant on the first assessment day (*M* = 2.46%, *SE* = 2.9%, $p = .396$) but was on the day 7 assessment (*M* = 6.06%, *SE* = 2.9%, $p = .0366$). Additionally, post-hoc tests indicated that the effect of day was not significant for the gesture plus translation condition (*M* = 5.3% , *SE* = 2.9%, $p = .067$), or the translation plus example condition (*M* = 1.7%, *SE* = 2.9%, $p = .556$).

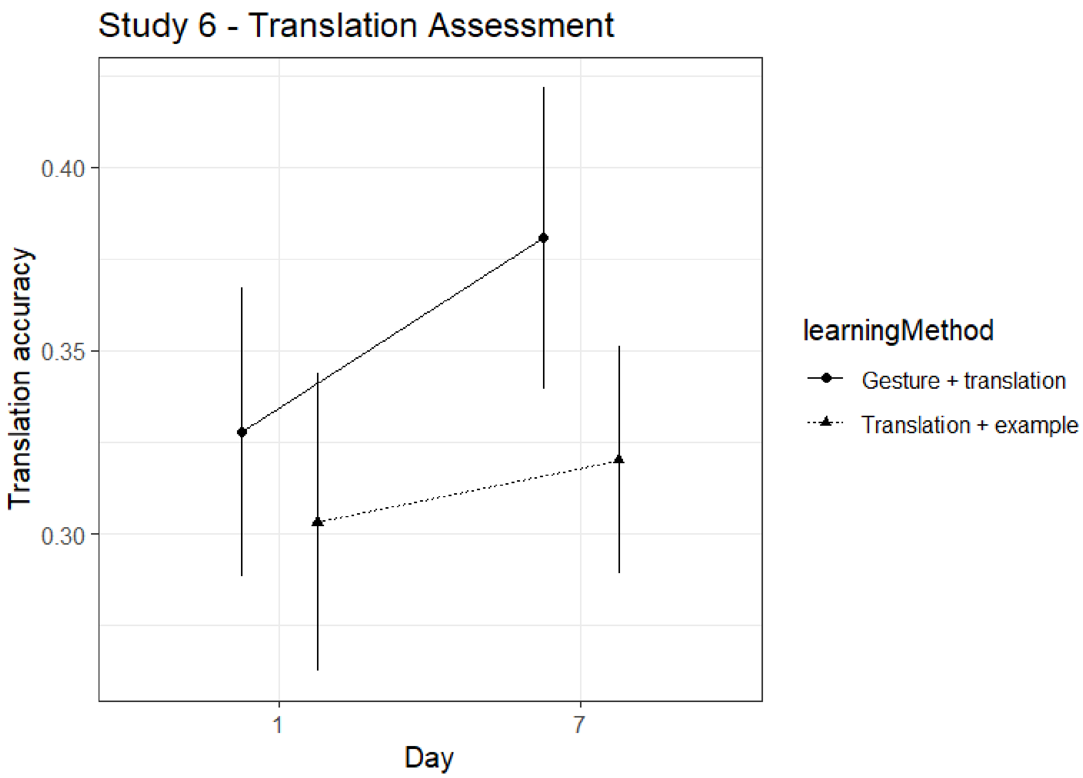


Figure 24. *Graph displaying the mean translation scores across assessment day 1 and 7 for words learnt through gesture plus translation cues and translation plus example cues. The within-subject error bars reflect the variability of the data using 95% confidence intervals.*

Relatedness Judgement task

Accuracy data

A one-way t-test was conducted in each learning condition to check that performance was above chance. For the gesture plus translation learning method, performance was greater than chance (*Mean accuracy* = .782), $t(42) = 14.66$, $p <.001$, Cohen's d = 2.24. Performance was also above chance for the translation plus example learning method ($M = .729$), $t(42) = 10.82$, $p <.001$, Cohen's d = 1.65.

A 2 (learning method) x 2 (day) x 2 (relatedness) ANOVA was conducted on the reaction time accuracy data (see Figure 25). The ANOVA found a significant main effect of learning method, $F(1,42) = 13.17$, $p <.001$, $\eta^2_g = .02$, with words learnt through gestures plus translations having greater accuracy (*Mean accuracy* = .781, $SE = .0194$) than words learnt through translation plus examples ($M = .729$, $SE = .0212$). There was also a significant effect of relatedness, $F(1,42) = 20.20$, $p <.001$, $\eta^2_g = .11$, indicating that unrelated trials had greater accuracy ($M = 0.818$, $SE = 0.0212$) than related trials ($M = 0.693$, $SE = 0.0257$). There was no significant effect of day, $F(1,42) = 3.35$, $p = .074$.

There was also a significant interaction between learning method and relatedness, $F(1,42) = 10.86$, $p = .002$, $\eta^2_g = .02$. Unrelated trials had greater accuracy across both learning methods (gesture plus translation $M = 0.819$, $SE = 0.0212$; translation plus example $M = 0.816$, $SE = 0.0238$) than related trials (gesture plus translation $M = 0.743$, $SE = 0.0235$; translation plus example $M = 0.643$, $SE = 0.0327$). There was no significant interaction between learning method and day $F(1,42) = 0.02$, $p = .894$, between day and relatedness $F(1,42) = 0.81$, $p = .375$, or between learning method, day and relatedness $F(1,42) = 1.06$, $p = .309$.

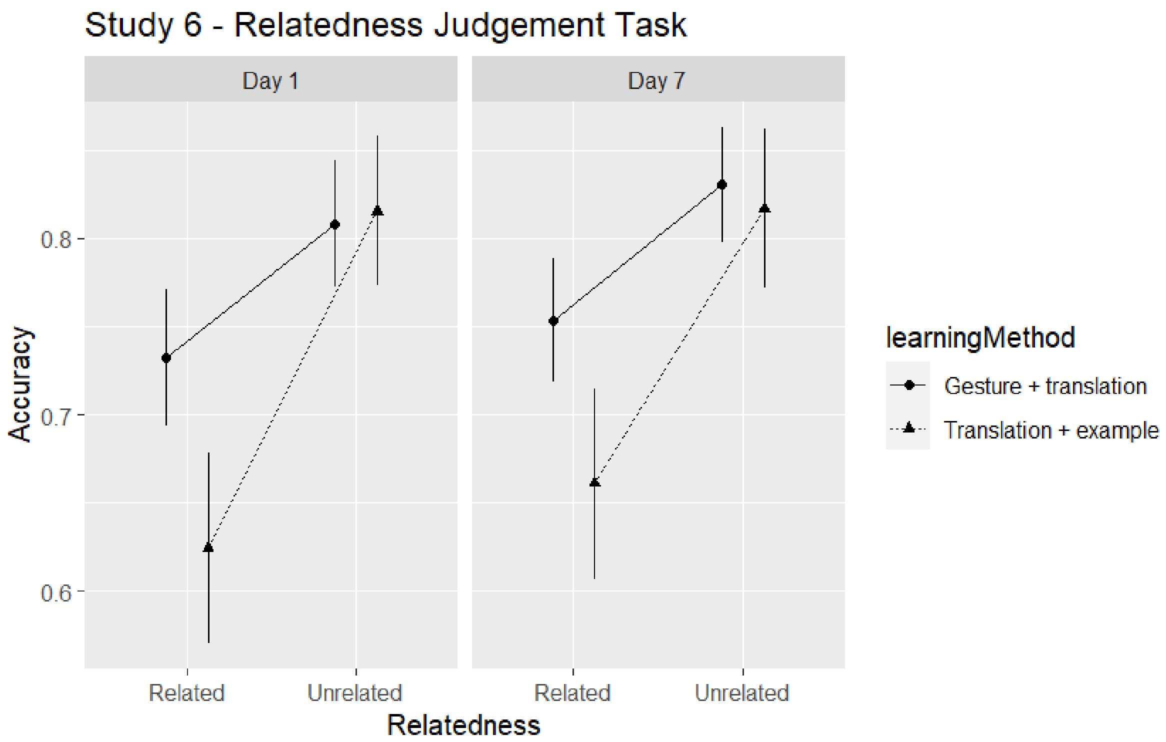


Figure 25. Graph displaying the accuracy for the Relatedness judgement task across the factors day, relatedness and learning method. Within-subject error bars reflect the 95% confidence intervals.

D prime data

To investigate the sensitivity of responses during the relatedness judgement task, participants' responses were categorised into hits, misses, false alarms and correct rejections (see Table 3). For day 1, correct rejections made up the most trials at 40.87%. Hits made up 33.26% of trials, misses 15.94% and false alarms 9.93% of trials. For day 7, again the majority of trials were correct rejections making up 41.20%. Hits made up 35.23% of trials, misses 14.62% and false alarms 8.95%.

Table 3

Visual display of how responses were recorded into sensitivity categories.

		Stimuli: Related	Stimuli: Unrelated
Response	Related	Hit	False alarm
	Unrelated	Miss	Correct rejection

From this, d-primes were calculated for participants across both days and split by learning method. For the day 1 assessment, larger d-primes were found in the gesture plus translation condition (*Mean d'* = 1.61, *min* = 0.10, *max* = 3.57) than for the translation plus example condition (*M* = 1.32, *min* = -0.32, *max* = 3.70), indicating that sensitivity to the stimuli was further from chance (d' = 0). For day 7, larger d-primes were again found in the gesture plus translation condition (*M* = 1.81, *min* = -0.16, *max* = 4.07) than in the translation plus example condition (*M* = 1.51, *min* = -0.44, *max* = 4.04). For a visual display of these results, see Figure 26.

A 2 (learning method) x 2 (day) ANOVA was conducted on the d-prime data (see Figure 27). There was a significant main effect of learning method, $F(1,42) = 8.63$, $p = .005$, $\eta^2_g = .02$, indicating larger d-primes for the gesture plus translation learning condition (*Mean d'* = 1.71, *SE* = 0.143) than for the translation plus example condition (*M* = 1.41, *SE* = .149).

There was also a significant main effect of day, $F(1,42) = 5.86$, $p = .020$, $\eta^2_g = .01$, indicating larger d-primes on day 7 (*M* = 1.66, *SE* = 0.155) than on day 1 (*M* = 1.46, *SE* = .0131). There was no significant interaction between learning method and day, $F(1,42) = 0.02$, $p = .901$.

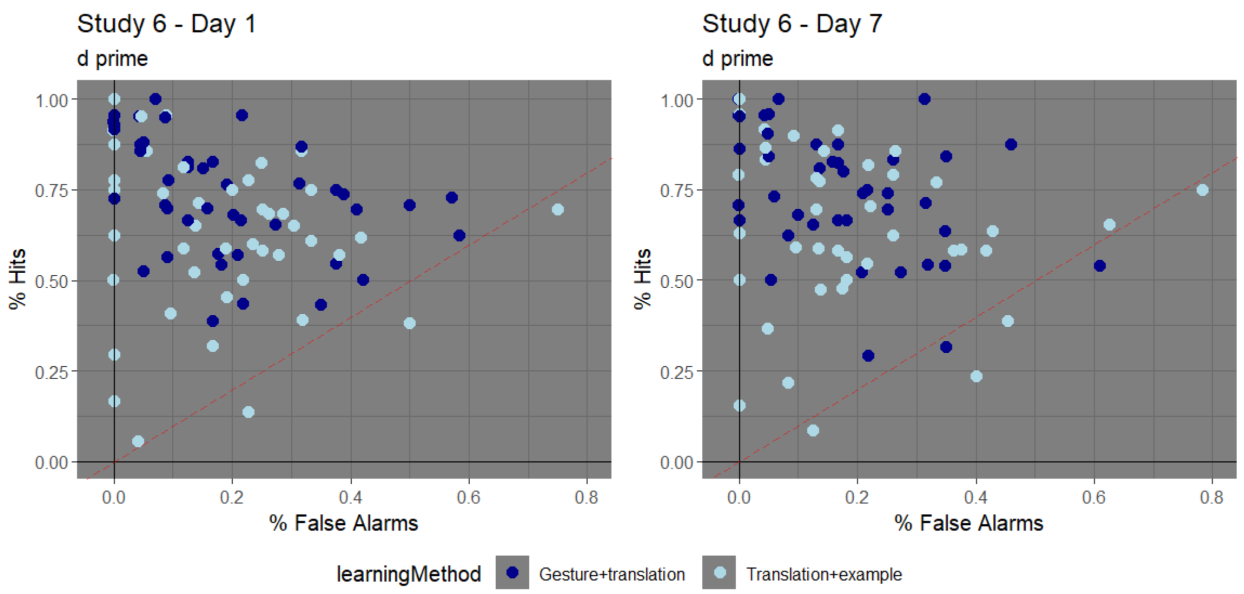


Figure 26. *Graphs showing the proportion of false alarms and hits for day 1 and day 7 assessments across the two learning methods. The red dashed line indicates a d prime of 0 (responses most closely associated with chance).*

Figure 27. *Graph displaying the d prime scores across assessment day 1 and 7 for words learnt through gesture plus translation cues and translation plus example cues. The within-subject error bars reflect the variability of the data using 95% confidence intervals.*

<u>Transparency and Openness</u>

The pre-registration stated that a 3-way ANOVA with learning method, day and relatedness would also be conducted on the d-prime data. However, hits and misses are not sensitivity categories within the unrelated trials and correct rejections and false alarms are not categories within the related trials. Thus, it is not possible to apply d-prime analysis to this data. Signal detection theory is not appropriate for such split data. It is also important to note that the d-primes calculated here were completed on data that had been slightly reduced due to reaction time outliers being removed (15.71% of trials). See Appendix H, p.182, for analysis of d-primes without reaction time outlier removals (the effect of learning method in the 2-way ANOVA remained significant, $p = .009$).

Reaction Time data

The reaction time data (filtered for reaction time outliers) for correct responses in the relatedness judgement task was analysed using a 2 (learning method) x 2 (relatedness) x 2 (day) ANOVA (see Figure 28). There was a significant effect of learning method $F(1,42) = 4.07$, $p = .050$, $\eta^2_g = .003$, indicating longer reaction times in the translation plus example condition ($M = 1199$, $SE = 29.5$) than in the gesture plus translation condition ($M = 1177$, $SE = 28.5$).

There was a significant effect of day $F(1,42) = 35.69$, $p < .001$, $\eta^2_g = .078$, with longer reaction times on day 1 ($M = 1249$, $SE = 25.8$), than on day 7 ($M = 1126$, $SE = 34.2$). There was a significant effect of relatedness $F(1,42) = 4.14$, $p = .048$, $\eta^2_g = .005$, indicating longer reaction times for related trials ($M = 1202$, $SE = 29.7$), than for unrelated trials ($M = 1174$, $SE = 29.0$). There was no significant interaction between learning method and day $F(1,42) = 0.32$, $p = .575$, between learning method and relatedness $F(1,42) = 2.75$, $p = .105$, between day and relatedness $F(1,42) = 0.52$, $p = .477$, or between learning method, day and relatedness $F(1,42) = 0.23$, $p = .633$.

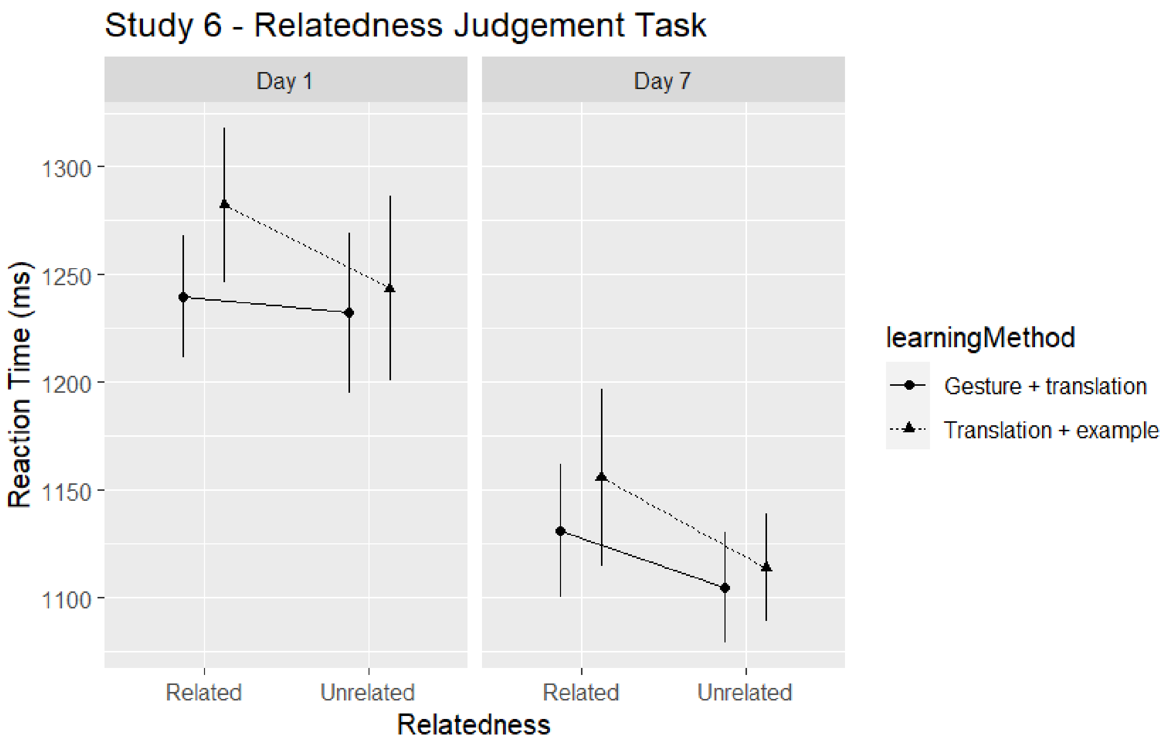


Figure 28. *Graph displaying the reaction times for the Relatedness judgement task across the factors day, relatedness and learning method. Within-subject error bars reflect the 95% confidence intervals.*

Discussion

Study 6 aimed to investigate whether the gestural advantage in early L2 word learning could also be evident in a speeded explicit task. The speeded relatedness judgement task found quicker and more accurate responses in the gesture-based learning condition than the non-gesture-based learning condition. Additionally, the gesture-based learning condition had greater d-primes, indicating that responses were further from chance, than the non-gesture-based learning condition. However, the gestural advantage was only evident in the explicit un-speeded translation task when the data was split by assessment day. Interestingly, and in contrast to the previous studies, participants in Study 6 displayed marked improvements in assessment day 7 compared to day 1. Possible explanations for these results are explored.

Relatedness Judgement task

The relatedness judgement task found evidence that the gestural advantage does prevail in a speeded task in early L2 learners. Words learnt through the gesture plus translation condition had significantly greater accuracy than the translation plus examples condition. Additionally, this greater accuracy was not at the expense of speed, with the gesture-based learning condition also producing quicker reaction times. This demonstrates that the beneficial effects of gestures are not confined to a controlled, explicit translation task.

L2 words learnt through gestures were able to be quickly and accurately semantically discerned. This suggests that gesture-based learning cues can increase deeper semantic understanding than translation-based learning cues. The learning conditions used were equal in the number of cues and the amount of disambiguating information they provided and only differed in the inclusion of a gesture cue or not. Therefore, the results support the privileged access account's involvement in the gesture advantage found in the L2 word learning literature. The accuracy data for this task also supports this account, with greater accuracy being recorded for L2 words taught through gestures plus translations than for words taught through translations plus examples. This gestural benefit was evident under speeded conditions when responses longer than 2000ms were removed.

The relatedness judgment task also found significant effects of relatedness. Unrelated trials had quicker reaction times than related. This is inconsistent with the previous literature and typical semantic priming effects both with L1 words (Kelly et al., 2015), and L2 words (Dittinger et al., 2016, 2019). A possible explanation for this difference in reaction speed will subsequently be discussed. Like the reaction time data, unrelated trials also had greater accuracy. Again, this does not align with typical semantic priming effects. However, Dittinger et al. (2016) found the same pattern in their semantically relatedness task involving L2 words and pictures. They explain their results through a potential response bias towards rejection. This is proposed to result from high task difficulty that elevates response uncertainty in participants, making them more likely to respond with 'unrelated' and thus having greater errors in related trials. Dittinger et al. (2016) did not conduct any further analysis on this response data and so could only suggest this response bias as an explanation.

Interestingly, the results found a significant interaction between relatedness and learning condition. Accuracy was greater for unrelated trials in both learning conditions, but this difference was larger for the translation plus example condition (see Figure 25). This suggests that words taught through translations plus examples had greater rejection response bias, which implies greater uncertainty for these words. This explanation could also be applied to the unexpected results in the reaction time data. If participants have a rejection response bias, then they are likely to be hesitant, thus taking longer, when they then respond with 'related', despite being confident in their decision.

Taking these potential response biases into account, the answers for the relatedness judgement task were categorised for their sensitivity based on signal detection theory

(Macmillan & Creelman, 2004). From this, d-primes were then calculated to inform of the consistency of responses and identify response patterns. Results found that the gesture plus translation condition had larger d primes (indicating responses were further from chance) than the translation plus example condition. This supports the findings from the accuracy data which suggested that participants' responses to words learnt through the translation plus example condition had greater rejection response bias and more uncertainty in their answers.

Translation task

When reviewing the results of the translation assessment, Study 6 was unable to replicate the previous gesture-based learning advantage found in the explicit task in Study 4. Although the translation task found the gesture plus translation learning condition to have descriptively greater translation accuracy than the translation plus example learning condition, the difference was not significant. However, the differences between the two learning methods did reach significance on the day 7 assessment.

The improvement across assessment days is an interesting effect that has not been present in any of the previous studies. Whereas Studies 1-5 all showed a decay in memory for the L2 words, Study 6 found that translation scores on day 7 were significantly higher than on the day 1 assessment. One explanation for these results could come from the additional exposure to the L2 words and concepts provided by the relatedness judgement task. Frequency of exposure to L2 words has been found to increase the integration of the newly learnt words (Magnuson et al., 2003). Although no additional learning was provided between the two sessions, the relatedness judgment task delivered on day 1, after the first translation task had been completed, may have allowed for continued learning during the task, as the actions were presented with correct L2 speech for half of the trials.

Although the same could be said to be possible for the previously used action-speech integration task, with 50% of trials containing congruent pairings, key differences between the task designs must be considered. In comparison to the implicit action-speech integration task, the relatedness judgment task was explicit. This would give participants far more awareness of the congruency of the stimuli as they were actively trying to judge this as part of the task. Within the action-speech integration task, the participants' task was to identify (under speeded conditions) the gender of the speech, and so it would be unlikely that they could pick up additional learning from the congruent action-speech pairings.

This improvement across days was also reflected in the outcome measures from the speeded explicit task. The relatedness judgement task recorded quicker reaction times for responses on day 7 than on day 1. This may just reflect practice effects, with participants becoming more familiar with the task and therefore perceptual processing increasing for the second assessment session (Hussain et al., 2009). However, it could also reflect greater confidence in responses and thus quicker reaction times on day 7. This is supported by the fact that participants had greater accuracy on the day 7 translation task than on the first assessment on day 1. Additionally, significant differences across days were also recorded for the d-prime scores. Larger d-primes were found for the day 7 assessment compared to the day 1 assessment. This again could be explained through additional learning and practice after the first assessment session. However, no differences across day were reflected in the accuracy data for the relatedness judgement task, suggesting that any additional learning gained may only have been available during un-speeded explicit retrieval.

Both learning conditions were equal in the number of cues provided, this is important because, as suggested by the DCT (Clark & Paivio, 1991), an uneven number of cues could have impacted learning. Additionally, both contained equal disambiguating information by containing the L1 translation text as well as a cue that provided context to the word. It is important to note that the additional example text given in the translation plus example learning condition was matched to the gesture videos and not the action videos. As a result, the contextual information across learning conditions was equal.

Therefore, the only differing variable between the two learning conditions was the use of a gesture. The privileged access account encompasses theories that propose gestures have specialist access to action representations and motor traces. When learning the L2 words, the gestures are automatically integrated with the speech, connecting these motor traces and action representations that have been activated to the L2. This activation of these systems is specific to gestures and independent from other visual modality pathways (Engelkamp, 1986). Due to the integration, during retrieval these traces and representations are activated again, aiding memory.

Overall, the pattern of results is consistent with the previous studies, with gesture-based learning methods producing greater translation scores than non-gesture-based learning. These results extend the previous literature by indicating that the gestural advantage is not limited to only affecting the controlled retrieval of L2 information but can also be beneficial on a deeper level in speeded semantic tasks. Since the number of cues and the level of disambiguating information provided was controlled across both learning conditions, the privileged access account appears to be able to explain this advantage. Gestures have the ability to provide unique access to motor traces that create a more memorable learning environment.

GENERAL DISCUSSION

The previous literature on the use of gesture cues in L2 vocabulary learning were limited in several key areas. Firstly, most of the previous studies have relied on explicit behavioural outcome measures, with few implementing implicit measures of learning. Additionally, few studies have presented gestures as the sole cue to an L2 word meaning, without accompanying L1 words. The studies presented in this thesis aimed to extend this previous work by measuring L2 learning through traditional explicit translation tasks, as well as speeded implicit and speeded explicit assessment measures, to determine if the use of gesture-based cues can facilitate semantic encoding of L2 words. The benefits of gesture cues in learning when presented in isolation was also explored. Additionally, the studies conducted aimed to investigate the mechanisms and factors that contribute to the gestural advantage displayed, when compared to other non-gestural cues.

Table 4

Overview of the studies conducted in this PhD

	Learning conditions	Cues	Store encoded into (DCT)	Rationale	Mechanism being tested	Outcome measures being tested	Results
Study 1 (online)	Gestures vs Translations	1 vs 1 Equal	1 NV vs 1 V Equal	Test gestures in isolation – are gestures still beneficial on their own?	Privileged access and disambiguation	Translation task Action-speech integration task	TT: Greater learning through G A-S: SC effect in G
Study 2 (online)	Gestures vs Translations and examples	1 vs 2 Unequal	1 NV vs 1 V Equal	Do gestures just provide additional context to the words?	Privileged access and disambiguation	Translation task Action-speech integration task	TT: Equivalent learning A-S: No SC effects
Study 3 (online)	Gestures vs Gestures and Translations	1 vs 2 Unequal	1 NV vs 2 NV+V Unequal	Does inclusion of translation further enhance advantage? Does G+T provide greater disambiguation?	Disambiguation	Translation task Action-speech integration task	TT: Greater learning through G+T A-S: SC effect in G+T on day 7
Study 4 (in-person)	Gestures and translations vs Translations and examples	2 vs 2 Equal	2 NV+V vs 1 V Unequal	Testing the involvement of privileged access when learning methods have equal levels of disambiguation and number of cues	Privileged access	Translation task Action-speech integration task	TT: Greater learning through G+T A-S: No SC effects
Study 5 (online)	Gestures	NA	NA	A more simplified study design that allows for greater learning may be required to display semantic congruency effects in newly learnt L2 words	NA	Action-speech integration task as an effective measure of sematic congruency	A-S: No SC effects
Study 6 (in-person)	Gestures and translations vs Translations and examples	2 vs 2 Equal	2 NV+V vs 1 V Unequal	Does the gesture advantage extend beyond an explicit translation task in early L2 learners? Is it still present in a speeded task?	Privileged access	Translation task Speeded, explicit Relatedness judgement task	TT: Greater learning through G+T on day 7 RJT: Greater accuracy, larger d-primes and quicker RTs through G+T

Note. NV = Nonverbal strore, V = Verbal store, G = Gesture learning condition, G+T = Gesture plus transaltion learning condition, TT = Translation task, A-S = Action-speech integration task, RJT = Relatedness judgment task, SC = semantic congruency.

Thesis overview

Study 1

Study 1 investigated the use of gestures as the sole cue to L2 language learning. Based on the RHM's predictions on lexical and conceptual links, the idea that gestures could fast-track L2 learning, without the accompanying L1 translation, was explored. Participants learnt half of the Mandarin action words through gesture cues (with limited exposure to the L1) and the other half through L1 translation cues. The study implemented a speeded, implicit assessment of L2 semantic integration, as well as an L2 to L1 translation task. Greater translation accuracy was recorded across both assessment days (1 day after learning and 7 days after learning) for words learnt through gestures compared to words learnt through translation cues. Additionally, words learnt through gestures displayed semantic congruency effects in the action-speech integration task on the first assessment day, with incongruent stimuli recording longer reaction times. This was interpreted as the semantics of the L2 words having been automatically integrated. No such semantic congruency effects were observed for words learnt through L1 translations, indicating that this deep semantic encoding was unique to the gesture condition.

Study 2

One possibility is that gestures led to better learning in Study 1 because they provide greater disambiguation on the context of the L2 action words than the L1 translation. While an action word presented in isolation (e.g., light) is semantically underspecified, the iconic gesture learning condition may disambiguate the L2 word by embedding it into a specific action context (e.g., lighting a match on the match box), resulting in better learning of the L2 word. Study 2 examined if this disambiguation account can explain the gestural advantage observed in Study 1. To provide the two learning conditions with equivalent contextual information, the L1 translations (e.g., open) were accompanied by example text (e.g., to open a book) that matched the semantic information provided by the gesture. No differences were found in the translation accuracy between learning conditions, indicating that the disambiguating, contextual information provided by gestures does play a role in the gestural advantage to some extent. Study 2 was not able to replicate the speeded implicit task findings from Study 1, with no semantic congruency effects being found. This raised questions of the stability of these implicit effects.

Study 3

Although Study 1 found that the gesture advantage does persist when gestures are the only cue to learning, it is not known whether the combination of gesture and L1 word would enhance or hinder learning, as compared to learning with gesture as the sole cue to L2 word meaning. Application of the RHM and levels of processing theories would predict that gestures in isolation are more effective than the combination of gesture and L1. However, application of the DCT would suggest that the L1 translation would act as an additional cue and thus provide greater encoding. It could also provide an even greater level of disambiguation. To answer this question, in Study 3 gestures presented in isolation were compared to a gesture plus L1 translation learning condition. Greater translation accuracy was recorded for the gesture plus translation learning condition, suggesting that the extra cue and level of disambiguating information aided learning. But again, the results in the action-speech integration task did not replicate the findings from Study 1, with only a semantic congruency effect in the expected direction appearing on day 7 for the gesture plus translation condition.

Study 4

One limitation of Studies 2 and 3 was that the number of cues and the degree of disambiguating information was not matched between the learning conditions (see Table 4). Study 4 therefore used learning conditions that were equal with respect to these variables so that the involvement of the privileged access mechanism could be directly tested. A gesture plus translation learning condition was compared to a translation plus example learning condition. Greater translation accuracy was recorded for words learnt through the gesture-based learning condition on both assessment days. This suggests that privileged access also contributes to the gestural advantage. However, again no semantic congruency effects were found for the speeded, implicit task.

Study 5

Considering the lack of a consistent effect found in the action-speech integration task across Studies 1-4, the ability of the task to measure semantic integration and the presence of such implicit effects in early L2 learning was questioned. One possibility is that this implicit task is simply not reliable as a measure of deep semantic integration in early stage L2 word learning. Another possibility is that the task is reliable in principle but only as long as the learning situation (including number of words in the learning set) is not overly complex. To create a more simplified study design and offer one last test of the task, Study 5 tested the use of gesture cues as the only learning method on just 12 Mandarin words. The study was unable to replicate the semantic congruency effects found in Study 1. The presence of such learning effects at a speeded and implicit level was questioned.

Study 6

This final study aimed to investigate if the gestural advantage during L2 word learning, found consistently in the explicit translation task, could also be evident in speeded settings. The lack of a consistent semantic congruency effect in the action-speech integration task suggests that any conceptual links from the L2 that might be developing may still be too fragile in these early stages of L2 acquisition to be evident under implicit, speeded conditions. The learning methods used in Study 4, gesture plus translation and translation plus example, were compared in Study 6 to further test the privileged access account of L2 learning within a new task. A relatedness judgement task was developed that required participants to make speeded, explicit decisions on the congruency of action videos and L2 auditory stimuli. Greater accuracy and d-primes (responses further from chance) were found for L2 words learnt though gestures plus translation in the speeded explicit task than for words learnt through translation plus examples. Additionally, quicker reaction times were also recorded for the gesture-based learning condition in this task. The results show that the benefits of learning through a gesture-based learning method can extend to a speeded task in early L2 learners. However, in the explicit translation task, greater accuracy for the gesture-based learning method was only evident on day 7. This suggests that the gesture advantage may be fragile and require a high level of learning and exposure.

Theoretical rationale: Revised Hierarchical Model

The theoretical background for these studies was based on the RHM (Kroll & Stewart, 1994). This model of translation production suggests that initially, when in the early stages of L2 learning, conceptual links between the L2 words and their meanings are underdeveloped. Instead, translations from the L2 are mediated by the strong lexical links. This longer route is facilitated until the direct conceptual links from the L2 are established. Research has suggested that these direct conceptual links may be able to develop sooner than previously thought- with evidence from less proficient learners suggesting they can, at least partially, access some conceptual information directly from L2 (Dufour & Kroll, 1995; Sunderman & Kroll, 2006).

Therefore, extralinguistic, meaningful cues could be used to facilitate the development of these direct conceptual links in very early L2 learners. Picture cues (Palmer & Havelka, 2010) and gesture cues (Macedonia & Von Kriegstein, 2012) have both been found to be beneficial in L2 word learning, with the latter displaying greater learning effects. Thus, gesture cues would provide the best opportunity to develop these conceptual links and display L2 semantic effects in early learners. But, to truly test if such cues would be able to develop theses conceptual links in the early stages of learning, as little access to the lexical links as possible during learning would need to be ensured. This would allow the greatest opportunity for the direct conceptual route to be taken and developed. As well as investigating if these conceptual links could be strengthened in early learning, the studies aimed to gain understanding on why gestures are advantageous over other cues. The mechanisms behind this gestural advantage were therefore explored.

Mechanisms proposed

Through the series of studies undertaken, I have aimed to test a number of different explanations that may contribute to the gestural advantage observed in the literature. The cues used were manipulated and paired together systematically in order to gain understanding of the involvement of the proposed mechanisms.

Dual Coding Theory

The Dual Coding Theory (Clark & Paivio, 1991) proposes that stimuli are encoded through verbal or nonverbal modality pathways into two separate, but interconnected, systems. Encoding through both modalities is believed to have additive effects for memory and recall, with greater retrieval being recorded for information coded through both visual pictures and verbal text (Paivio & Csapo, 1973). The DCT can therefore explain the advantages for learning involving a combination of any verbal and extralinguistic cues. The theory can provide an explanation for the gesture learning effects found in the past literature, when gestures were used as an additional cue alongside L1 text literature (e.g., Kelly et al., 2009; Macedonia et al., 2011), as well as the results of Study 3 in this thesis. The benefit of dual coding through two modalities was evident in Study 3 when the use of gestures as the sole cue in learning was compared to a gesture plus translation learning method. Greater translation accuracy for the L2 words was evident when the two cues provided encoding through both verbal and extralinguistic cues (gesture plus translation).

But, as well as the symbolic modalities (verbal and nonverbal), Paivio (1978) also identified the different sensory modalities that stimuli can be presented in, including visual, auditory, tactual and kinaesthetic/motor. Some extralinguistic stimuli, such as actions, are believed to be encoded through multiple modalities, with Clark and Paivio (1991) stating that "nonverbal components of the motor system include both kinaesthetic and visual images" (p.185). The involvement of a motor route to encoding has been explored in the literature, with suggestions that this route may provide greater learning than a visual-only route. Huang et al. (2019) interpret the DCT to predict an 'incremental involvement' of sensory modalities: 'auditory < auditory + visual < auditory + visual + kinetic/motor-manual' (p.189). Therefore, compared to picture cues, gestures would also be encoded through the motor route and that this could explain the greater learning compared to other extralinguistic cues. The DCT does not appear to specify if, in the context of cues to language learning, gestures must be performed to be encoded through the kinetic/motor modality. However, as their study did not involve performing the gestures, Huang and colleagues (2019) propose that enactment is not necessary to elicit motor encoding and that simply observing such movement, particularly within a 'focused training context', is sufficient to tigger a 'mental simulation of that gesture' (p.189).

This proposal, that the kinaesthetic route may be taken without movement from the learner, has also been supported by educational research. Kassim (2018) compared new L2 vocabulary learning (nouns, verbs and adjectives) within text through translation plus static image glosses and translation plus animated image glosses. Learning was tested through a 6-AFC task. Greater translation accuracy was recorded for the translation plus animated image learning condition- both immediately after learning and one week later. These benefits have also been demonstrated with video clips over picture cues in dual coding (Al-Seghayer, 2001). This research supports Huang et al.'s (2019) interpretation of the DCT, wherein dual coding that also includes the motor modality has additive effects on memory and recall over dual coding that does not encode via this route.

Although not specified within the DCT, the use of multiple cues within learning was also explored in this series of studies. Study 2 used learning conditions (gesture compared to translation plus example) which are both encoded into a single modality store (nonverbal and verbal, respectively). The two learning conditions provided an unequal number of cues to the L2 word (1 cue vs 2 cues). Despite this, a lack of difference in the L1 translation recall was found. Taken in conjunction with the results of Study 1 (gesture > translation), these results suggest that multiple cues during learning is beneficial to memory. Additionally, the results of Study 2 show that, despite having less cues, gestures presented in isolation were still equivalent to a ‘two-cue’ condition. This demonstrates the robust benefits gestures have for L2 word learning. However, Huang et al.'s (2019) interpretation of the incremental involvement of sensory modalities of the DCT would predict gestures to be encoded through both the visual and motor pathways. This would suggest that gestures have more routes to encoding and so should provide greater recall than the translation and example learning condition, which only provides verbal store encoding through the visual route. Another factor or mechanism must also be involved that can explain these results.

The DCT could also be applied to further understand studies that compared gesture plus translation with translation plus example text (Studies 4 and 6). Although these two learning methods provided an equal number of cues, the gesture-based learning method still resulted in superior learning. This can be interpreted on the basis of the DCT. Whereas the translation-based condition only provided cues to the verbal system store, the gesture-based learning condition allowed for encoding into both the verbal and nonverbal store. Additionally, gestures may also be providing a third, motor route to encoding providing a further advantage to recall memory. The DCT could therefore, alongside other explanations, be used to explain the advantage found in the gesture plus translation learning condition used in these studies.

Therefore, despite the claim that the learning conditions in Study 4 and 6 were equal in the number of cues and disambiguating information, application of the DCT would suggest that they were not equal in the number of encoding routes. In order to test other mechanisms (such as the privileged access account) more directly, the gesture plus translation learning condition should be compared to a learning condition that also provides encoding into both the verbal and nonverbal store, such as a picture plus translation condition.

However, there were several practical reasons why L1 translation were originally selected as the other cue. One main reason for the decision to use L1 translation cues relates to the design of the speeded measures. The action-speech integration task and the relatedness judgment task both used short videos of actions being performed and therefore objects were displayed in the clips. Picture cues for the L2 action words would also need to include the objects to provide clear information on what the action is. However, this would then mean that L2 words learnt via picture cues would have an advantage over gestures learnt words in these speeded tasks as the objects used in the action videos would not be novel. This familiarity with the object stimuli would likely provide a benefit.

Disambiguation account

The disambiguating properties of gestures were also explored as an explanation for the gestural advantage. Gestures can be helpful tools to disambiguate speech when homophones are used (Holle & Gunter, 2007), in poor listening conditions when the probability of mishearing speech is high (Obermeier et al., 2012) and when early L2 learners try to disambiguate homophones in their non-native language (Ray, 2015). Research has investigated the use of gestures to help communicate and disambiguate L2 speech (Gullberg, 2006), but little research has looked into the specific disambiguating properties gestures can provide to L2 word learning. Considering the benefits of their contextual information in L1 and L2 speech, similar effects were presumed to occur when used in an L2 word learning setting. This is particularly apparent given that in this context, and across much of the literature in the area, L2 words are often semantically underspecified. For example, when presented on its own, the action verb 'cut' can be ambiguous, as it can be used in several different contexts (e.g., to cut your knee, to cut the grass, to cut paper with scissors). The presentation of a gesture (or even example text) could disambiguate such a word by providing semantic specification (e.g., index finger and middle finger outstretched in a 'v' shape to demonstrate scissors cutting), which may in turn lead to greater learning.

After Study 1 demonstrated that the gesture advantage in L2 word learning persisted even when gestures were the sole cue to the meaning of the L2 word, it was still not clear what mechanisms were driving the effect. Translations were presented in isolation in Study 1 and therefore could not provide the specific contextual information akin to that found through gestures. When example text that provided similar contextual information as gesture was added to the L1 translation cue in Study 2, the gestural advantage disappeared. This provides support for the involvement of disambiguation in L2 word learning.

Following this, Study 3 observed that the efficiency of gesture-based learning could be boosted further by adding translation text alongside the gesture cues. As explored previously, this benefit could be explained through the DCT as the additional translation text would be encoded into the verbal store. But, as well as this, the additional cue may also be providing an extra level of disambiguation. The gesture cue provides disambiguation in the form of context of the word, whilst the additional L1 text provides specific focus and certainty on the exact lexical translation of the word.

Privileged access

Many theories (Engelkamp & Zimmer, 1985; Macedonia & Mueller, 2016; Macedonia & Von Kriegstein, 2012) have been developed to explain the gestural advantage in terms of the special qualities and access gestures possess. One key feature of gestures highlighted in this thesis is the assumption that the integration of gestures and speech occurs automatically during comprehension (Kelly, Özyürek & Maris, 2010). Although this integrated-systems hypothesis is not so much a theory for the gesture advantage, it does demonstrate the strong connections co-speech gestures have with spoken language and thus show how this integration could also be beneficial when gestures accompany L2 speech.

Throughout this thesis, I provided insight into a few of these key theories highlighted by the literature, including embodied representation (Macedonia & Repetto, 2017), self-involvement (Helstrup, 1987), enactment and motor trace theory (Engelkamp & Zimmer, 1985). The term privileged access was coined to encompass all theories that propose gestures to hold specialist access to motor traces and action representations. The aim of the series of studies presented here was not to determine which of these theories appears most prominent, but to determine whether the gestural advantage in L2 word learning can (at least partially) be attributed to privileged access. In particular, I aimed to see whether there is evidence for privileged access under automatic conditions in early L2 learners.

To investigate the role and strength of the privileged access theory, it was important to test if the gestural advantage would persist when other potential factors involved in the gestural advantage were absent. The potential involvement of privileged access was observed in Studies 1 and 2 as gestures were compared to a non-gesture-based learning method.

But these studies could not determine the unique role of privileged access as other factors, including the number of cues involved in learning and the level of disambiguation, were not equal across learning conditions. Studies 4 and 6 were designed to eliminate these other third variables and could therefore test the involvement of the privileged access mechanisms in isolation. In these studies, the gesture-based learning condition displayed greater L2 translation accuracy, as well as quicker reaction times and greater accuracy in the speeded explicit task in Study 6. This suggests that, compared to a translation-based learning method, gestures provided unique motor-traces or action representations that became activated when the L2 word and gesture were paired together. These motor traces are independent from other visual modality pathways that may become active with other extralinguistic cues (Engelkamp, 1986). During retrieval, these motor cortices become active again, as evidenced from neuroimaging studies (Macedonia et al., 2011; Mayer et al., 2015).

One of the key differences between the motor trace theory and the DCT is whether, for the advantage of gestures to prevail, the gesture needs to be performed or not. The motor trace theory was developed largely based on the enactment effect (Engelkamp, 1986). This effect draws on data that demonstrates greater memory for words and phrases that have been performed, with recent work supporting this theory (Macedonia, 2014; Macedonia & Mueller, 2016; Macedonia & Repetto, 2017). Therefore, the motor trace theory suggests that, for these specialised motor traces to become activated during learning, the gesture must be enacted by the learner. However, the DCT does not specify the need for performance of the gesture for encoding into the nonverbal store through the motor (kinaesthetic) modality. This suggests that just viewing the gestures being performed is enough to elicit benefits to memory.

As the majority of the studies completed in this thesis were conducted online, it was difficult to have control over the enactment of the gestures carried out. The experiment instructions told participants to "copy what the teachers say and do". However, this instruction was not repeated after the initial introduction and feedback from participants showed varying amounts of gesture performance (see Appendix I, p.184, for the self-reported enactment data from the end of study surveys for Studies 2, 4-6). As a result, it is difficult to draw conclusions from this research if enactment is necessary for the gestural advantage in L2 word learning to persist.

Recent research has directly assessed the differences. García-Gámez et al. (2021) investigated the differences in learning words of an artificial language (Vimmi) through either a 'seeing' learning condition or a 'doing' learning condition. Oral translation accuracy was recorded. Although greater recall was observed in the 'do' group, there was no statistically significant effect of teaching group when analysed by participant. The enactment group displayed less interference effects than the observing group when processing incongruent gestures. Additionally, the retrieval of L2 words was faster for the performance teaching group. Nonetheless, the use of congruent gestures for learning still produced significantly greater recall scores over any of the other teaching conditions for the 'seeing' group. This demonstrates that, although producing the gestures may be favourable to further enhance the advantage, re-enactment is not necessary for the gestural benefits to L2 word learning.

Despite the gesture advantage in L2 learning appearing to persist when only viewing the gesture and not enacting (Huang et al., 2019; Kelly et al., 2009; So et al., 2012), the motor trace theory focuses on the necessity of performance to fulfil the motor component of the enactment effect. However, the need for performance in order to engage the motor system has been questioned (Ianì & Bucciarelli, 2017). Many L1 studies have displayed equal recall for words and phrases practised with gesture/action performed tasks (subject-performed tasks, SPT) and gesture/action observed tasks (experimenter-performed tasks, EPT) (e.g., Cohen, 1981). Cohen (1989) explains this by stating that "enactment, either by the subject or by the experimenter, adds a dimension to the memory trace that facilitates retrieval. The actual pattern of the motor activity appears to be unimportant; the main thing is that some appropriate activity is involved." (p.72). However, the motor-trace theory discusses the enactment effect only in terms of self-performed tasks. This seems limited given the evidence for gesture observation also aiding recall.

Indeed, this research has been supported by evidence for the involvement of the motor system during experimenter-performed tasks. Completing an unrelated motor task during gesture observation, in which the same effectors are engaged as the experimenter, causes the gestural advantage to disappear (Ianì & Bucciarelli, 2017). However, when the participants' motor system was loaded with a task that involved different effectors (e.g., legs and feet) to the observed gesture (e.g., using arms), the benefits of gestures to recall remained. This suggests that the same motor areas are activated when observing a gesture as when performing (Buccino et al., 2004). This is further corroborated by research on mirror neurons (Rizzolatti & Craighero, 2004). The same motor components have been observed to be activated when observing actions as when actions are self-performed (Fadiga & Craighero, 2003). Considering the similarities across actions and gestures, this would suggest that this activation would also be present when just observing gestures. This research further questions the motor trace theory's focus on only the self-performance of gestures to aid memory.

The few studies that have observed advantages of self-performance (Engelkamp & Zimmer, 1997) have been explained through the item number and order of encoding trials (intermixed or blocked across EPT and SPT) (Engelkamp & Dehn, 2000). However, recent research from Sivashankar et al. (2023) controlled for such item order but still recorded greater benefits for self-performed gestures as opposed to viewing experimenter-performed gestures on a screen in L1 verb recall. The authors suggested that the physical presence of the performer is key in the observation group to engage attention and the mirror-neurone system.

To sum up, based on the research discussed, the observation of performed gestures/actions appears to provide an advantage for recall, with the same activation of motor areas as for self-performance. The research suggests that the gesture benefit to L2 word learning may be slightly greater when self-performance is involved. However, due to the motor trace theory's clear specification on the self-performance of enactment for the activation of motor pathways, the theory does not appear to be able to fully explain the results found in this thesis, due to the inconsistencies in self-performance. An adapted model that encompasses the evidence for the benefits of observing gestures and the activation of the motor system would be more fitting for this research.

Therefore, such evidence, along with the research conducted in this series of studies, supports the idea that multiple mechanisms may play a role in the gestural advantage and do not act exclusively from each other. In this way, the privileged access account was proposed to reflect this potential co-existence of theories that underpin the extensive gestural advantage in word learning.

Gesture plus translation learning methods in L2 teaching

Studies 4 and 6 of this thesis suggested that the most effective way to present gestures in L2 word learning is to include the accompanying L1 translation. This contradicts the initial application I proposed for the RHM theory- that gestures presented alone would allow for direct conceptual links from L2 to develop, thus eliciting greater learning. However, as discussed in the Introduction to Study 3 (p.78), there are several possible explanations for this, including the use of multiple cues, encoding into both the verbal and nonverbal store (DCT) and greater disambiguation.

Nevertheless, the practical advantages of using a gesture plus translation learning method can also be explored by applying other aspects of the RHM. The model proposes that early L2 word learning relies on development of lexical links between the L2 and the L1. As L2 proficiency increases, direct conceptual links from the L2 begin to form and strengthen, facilitating this route. A learning method that can facilitate both routes to compliment natural learning, whilst encouraging semantic learning, could be beneficial to ensure that conceptual links begin to develop. This is important as research suggests that even in fairly advanced learners, this shift to using established conceptual links does not always occur. Prince (1996) tested the learning and recall of L2 words through translation and conceptual methods in more advanced and less advanced L2 English speakers. Participants were French university students in which English was a compulsory subject. All participants had been studying English for 5-8 years, and the weak and advanced L2 group split was determined by performance on TOEFL. Half of the participants learnt the new L2 words through translation learning and the other half through context learning (the unknown words were given in L2 sentences). Shortly after the learning phase, translation recall and contextual recall tasks were completed by all participants.

Weaker bilinguals performed better compared to more advanced bilinguals when no transfer of knowledge was needed (translation learning with translation recall). In comparison, the more advanced bilinguals were better able to apply words learnt through translations to appropriate L2 context. This demonstrates a limit in the weaker learners' ability to apply L2 words to context, despite strong translation learning. Application of the RHM would suggest this reflects strong lexical connections between L1 and L2, but underdeveloped conceptual links, despite participants being quite far along in their L2 learning. Indeed, Prince states that these participants are "overdependent upon translation links and so have failed to develop certain processing strategies crucial to the effective use of context" (p.486).

Applying this interpretation to my findings suggests that in very early L2 learning, a gesture plus translation learning condition may allow for the learning through the natural lexical links, whilst still providing input (and potentially early development) of the conceptual connections. Such a learning condition may be beneficial to ensure that as the learner develops, they do not become overdependent upon the lexical links but instead can seamlessly begin to mediate translations through the conceptual links that have been developing through the exposure to concepts from the onset of learning. Prince (1996) highlights the importance of a learning strategy that can combine the effectiveness of both translation and contextual learning, which gesture plus translation learning cues may well do.

Teaching of L2 vocabulary should aim to introduce contextual information alongside L1 translations as early as possible. Not only could this learning method be useful for very early L2 learners (as displayed in Studies 4 and 6) but also for later learners, to prevent a heavy reliance on lexical links continuing beyond beginner stages of learning and allowing the development and access of conceptual links from the L2.

Limitations

The action-speech integration task

The action-speech integration task was used to measure the automatic semantic integration of L2 words with meaning across Studies 1-5. The design of this assessment was based on the Stroop-like task developed by Kelly, Creigh and Bartolotti (2010) to demonstrate the automatic integration of gestures and speech. The task was adapted to use actions instead, which have been found to also hold similar automatic-like processing (Kelly et al., 2015).

The task was first piloted with L1 English speakers and as expected, semantic congruency effects were found. Next, the task was tested on advanced L2 English speakers and semantic integration of the L2 words under these speeded, implicit conditions was again found. That is, trials in which the English speech and actions being performed did not match had longer reaction times than congruent trials. This supports the RHM's prediction that strong conceptual links from the L2 are developed in advanced learners.

However, despite finding semantic congruency effects in the action-speech integration task in native and non-native fluent speakers, no consistent effects were found for newly learnt L2 words. Study 1 provided promising results for the use of the task to measure the semantic encoding of L2 words taught through gestures, with semantic congruency effects emerging for words taught through this learning method, but not translation cues. Yet no other semantic effects were recorded in the subsequent studies apart from in Study 3 where an effect for the gesture plus translation learning method was recorded for the day 7 assessment. However, the gesture-only learning condition in this study appeared to display an opposite semantic congruency effect (semantically congruent trials recording longer reaction times than incongruent) to that found in the same condition in Study 1.

It was proposed that the conflicting findings in this task may be a result of the weak and fragile semantic congruency effects that are being displayed at these early stages of L2 word learning. An increase in the levels of learning of the L2 words (displayed through the explicit task) was also thought to help increase the chances of recording the semantic effects. Additionally, it was believed that interference from the other learning conditions, and the cues used, could be impacting the strength and consistency of the semantic effects displayed. However, even under optimal learning conditions in Study 5 (reducing the number of Mandarin words to be learnt, only using a gesture learning condition), still no semantic congruency effects were found.

The series of results from Studies 1-5 suggest that implicit semantic effects are not developed enough in the very early stages of L2 learning to be detectable in such a speeded task. The previous literature that has provided evidence for semantic connections in lower-level proficiency L2 learners did not investigate these effects under such implicit conditions. Sunderman and Kroll (2006) found meaning related interference in an explicit translation recognition task which had a speeded element (responses over 3000ms excluded), and (Dufour & Kroll, 1995) used an explicit categorisation response task to measure access to conceptual information (responses over 4000ms removed). Neither of these studies therefore tested under implicit measures or under speeded conditions as time demanding as in the action-speech integration task (outliers over 2000ms removed). In this way, this task was a novel measure for the automatic integration of newly learnt words, but perhaps it was too ambitious to expect to observe such effects under these conditions in early L2.

However, the task was a successful measure when used with non-native L2 speakers. The pre-test study for the action speech integration task used participants with high proficiency levels. On a 10-point scale, all participants had a self-reported English proficiency score of 5 or above for reading, understanding/listening and speaking. The mean scores for each of the categories was above 8 (see Appendix J, p.190, for full list of results). Just over half of the participants used in this study (n=23) were recruited through the University of Hull's SONA system and were students or staff at the university. The other participants (n=17) were recruited through Prolific which screened participants to ensure that their first language was not English, but they had listed English as their fluent other language. Therefore, given these participants were highly proficient, finding semantic congruency effects is not surprising and supports the previous literature (Talamas et al., 1999) as well as the RHM theory- that advanced L2 learners have developed conceptual links from the L2. Conducting the action-speech integration task on lower proficiency learners would be helpful to identify when L2 conceptual links become developed enough for automatic-like semantic congruency effects to be displayed.

It is also important to note that weak and inconsistent semantic congruency effects observed in Studies 1–5 could be the result of using action for this task instead of gestures, as in Kelly, Creigh and Bartolotti (2010). The comparison of speech-action and speech-gesture pairings using a similar task found that although the congruency between audio and visual information affected reaction times, the difference was greater for the gesture pairings (Kelly et al., 2015). This suggests that any semantic congruency effects found in the action-speech integration task used in Studies 1-5 are likely to be weaker than if gesture-speech integration had been measured. This is supported by the fact that, although a significant semantic congruency effect was recorded in both pre-test studies using the action-speech integration task, the differences in mean reaction times in my own studies were smaller than in Kelly, Creigh and Bartolotti's (2010) experiment using gesture-speech integration (see Appendix A, p.164, for the pre-test results).

Similar motor-traces for gestures and actions

The speeded assessment measures used in this series of experiments all involved action video clips as a means of assessing semantic learning. Due to the nature of iconic gestures, the gestures used during learning were often pantomime gestures that displayed the same movement pattern to those being presented through these actions (Hostetter & Alibali, 2008). For example, the gestures for pour, type and whisk have perceptual similarities and motor traces to the actions. On the other hand, there were other gestures which contained less perceptual similarities as they involved body-part gestures that used movements of the hands to act as an object, e.g., cut (hand used as scissors), open (hands open like a book) and unlock (hands act as key and lock). However, most of the gestures used in these studies contained the same or similar movement patterns to the actions. This could therefore suggest an advantage for gestures in the assessment measure using these action videos.

Although consistent semantic congruency effects were not observed in this series of studies, this is something important to consider if a task of this nature is to be applied to future research. Adaptation of the stimuli used in the task could help to ensure that the gestural advantage found is a result of greater conceptual understanding and connections, and not just from recognising a similar or unsimilar perceptual movement to the one learnt alongside the L2 word. This will be discussed further when I give suggestions for future work.

This potential effect of the similarity in motor perception between gestures and action could also be applied to the relatedness judgment task implemented in Study 6. As the L2 words were always paired with another action that had been learnt, the gesture-learnt words may have had an advantage in also being able to apply motor similarities to make a quick judgement and not only rely on the semantic information.

This highlights another issue with the relatedness judgment task, as well as the fact that no foils were included (i.e. there were no action videos related to words that had not been learnt), Study 6 also only presented the same action words (presented through L2 or action video) as a pair. For example, the unrelated trial for the word mǒ (meaning spread) would have an action video of lighting a match, and vice versa, the unrelated trial for diǎn rán (meaning light) would have an action video of spreading butter on bread. This was originally done to provide consistency with the pairs used in the action-speech integration task. However, the relatedness judgement task was an explicit task, meaning that the word pairs would have been more obvious to participants in Study 6 than in Studies 1-5. With the benefit of hindsight, the same pairings should not have been applied. Instead, different trial pairs, including video clips of actions not learnt, should have been used. The consistent pairing of words may also explain the marked improvements displayed in the day 7 translation assessment. With the L2 being paired with the related action video for half of the trials and one other cue for all other trials, additional learning was likely to have taken place.

Does this series of studies reflect normal L2 learning?

It must also be noted that the studies outlined in this thesis are limited in what can be inferred about standard L2 learning from the one-time learning session that was implemented in these studies (Gullberg et al., 2012). The results of such studies must be applied with caution to actual L2 learning, where a number of external factors can influence the effectiveness of a certain learning method on individuals. This concern has been highlighted by Kroll et al. (2010) who suggested that learning a small set of words in a new language in highly practised conditions “are likely to produce semantic effects that may be unrepresentative of actual L2 learning.” (p.5). It is important that this research is considered alongside naturalistic research (McCafferty & Stam, 2009), which can often reflect real life teaching practices more accurately. Indeed, the Total Physical Response technique that Asher (1969) developed and tested in his empirical work has been successfully incorporated into an educational setting, and forms the basis of current teaching at the Polis Institute (Polis, 2024).

Implications for future research

The results of these studies discussed in this thesis have presented opportunities for future research. Firstly, as mentioned previously in the section 'The action-speech integration task' (p.136), it would be useful to test the action-speech integration task in a low L2 proficiency population to determine if semantic congruency effects under speeded, implicit conditions can be evident at this stage of L2 learning. Such research would aid determining when the semantic connections from L2 become developed and strong enough to operate under these taxing conditions. If semantic congruency effects can be demonstrated in low level L2 learners, future research could then look to assess the use of gesture-based learning cues to help develop the conceptual links after several learning sessions, as opposed to just the one used in these studies.

Additionally, if future research plans to apply either of the speeded tasks used in this series of studies, it would be useful to determine if gesture-based learning can be generalised beyond the concepts presented through gestures. Future research using action videos that portray the use of the action in a related, but not the same, context as the gesture could be beneficial for several reasons. Firstly, as mentioned previously in the section 'Similar motor-traces for gestures and actions' (p.138), the gesture and actions used in this series of studies often contained the same, or at least very similar, movement patterns (Hostetter & Alibali, 2008). These low-level perceptual similarities between the gestures and actions could be argued to give words learnt through gestures an advantage through perceptual memory, as opposed to reflecting greater semantic learning. Thus, the use of new action videos, that have different movement patterns, in the assessment tasks would allow the semantic learning to be measured.

Another reason for exploring this generalisation stems from the disambiguation account – one of the possible explanations explored in this research for the gestural advantage. Despite the beneficial disambiguating properties of iconic gestures that have been highlighted, there may also be some limitations in using these cues for L2 word learning. Considering the highly semantic nature of iconic gestures (Beattie & Shovelton, 1999; McNeill, 1992), and their ability to provide specific, unambiguous information (Holle & Gunter, 2007), it may be questioned if, when used in word learning, they might be too specific when trying to apply words learnt to other contexts. Cross-cultural research using Dutch and Mexican participants shows that acting gestures (pantomime gestures) are the most frequently used gestures to depict verbs involving an object (Ortega & Özyürek, 2020). Additionally, this type of gesture was the most accurately guessed when presented to another group. This demonstrates the clear, universal understanding provided through gestures and begs the question whether having such narrow meaning applied limits the generalisation to other contexts. Even so, research has suggested that gestures can be useful for generalising information, with gestures produced alongside speech aiding the retention and generalisation of learning techniques to solve maths problems (Congdon et al., 2017; Singer & Goldin-Meadow, 2005). However, this research used deictic gestures, not iconic gestures (which hold specific semantic meaning), and thus may be more difficult to generalise learning from them in the context of L2 learning (although little research has been conducted in this area).

In order to test this, the action videos should vary in context from the gestures. For example, the action video used in this study for 'zip up' involved zipping up a jacket, the same as the gesture. But if the action used in the relatedness judgment task was slightly different, such as zipping up a bag, and participants still displayed learning-method effects in accuracy and reaction time, this would indicate an ability to generalise the gestural learning method. Therefore, assessment tasks which use related actions videos which can measure the generalisation of conceptual knowledge, would be helpful to ensure learning through iconic gestures can be applied to other contexts and are not limit in their application to other settings.

Conclusions

Overall, this series of studies investigated the gestural advantage in L2 word learning beyond the existing literature. Although this benefit of gestures still prevailed when presented as cues on their own, even better translation performance was observed when gestures were accompanied by the L1 translation. Greater semantic learning was also observed for these cues under speeded explicit conditions, demonstrating that the gestural advantage is not limited to slow, controlled retrieval, but is also evident under more automatic-like processing. The studies identified the involvement of several factors and mechanisms facilitating the gestural advantage, namely, the use of multiple cues, multiple encoding modalities (DCT), disambiguation and privileged access.

This research also explored the application of a speeded implicit task as an outcome variable to measure semantic integration of L2 vocabulary. More research is required to draw conclusions of when semantic effects, under such automatic conditions, can begin to be displayed in L2 learners.

Finally, this series of studies clearly demonstrates the benefits of integrating gestures into learning for L2 acquisition, which can have many implications for L2 teaching. Within the classroom, the use of gestures (through observing a teacher perform them or from self-performance) should be included during the initial introduction of new vocabulary (e.g., during the introduction to a new topic). The L2 words should be spoken whilst the gesture is performed, to enhance the integration between the gestures and speech. As demonstrated from this series of studies, the L1 translations should also be presented alongside the gestures.

As practice on this L2 topic increases, exposure to the gestures should continue to be integrated into learning to provide opportunity for direct conceptual links to develop. Additionally, it may be beneficial to incorporate these gestures into tasks to assess learners recall, for example the teacher performs the gesture and the students must respond with the L2 translation. This would allow for the development of the weaker and more effortful links between concepts and L2 (Kroll et al., 2010). Also, this may prevent a reliance on the stronger lexical links by encouraging students to practice using these weaker conceptual connections early on during learning (Prince, 1996).

REFERENCES

Al-khresheh, M. (2010). Interlingual interference in the English language word order structure of Jordanian EFL learners. *European Journal of Social Sciences*, *16*(1), 106–113.

Allen, L. Q. (1995). The Effects of Emblematic Gestures on the Development and Access of Mental Representations of French Expressions. *The Modern Language Journal*, *79*(4), 521–529. https://doi.org/10.2307/330004

Al-Seghayer, K. (2001). The Effect of Multimedia Annotation Modes on L2 Vocabulary Acquisition: A Comparative Study. *Language Learning & Technology*, *5*(1), 202–232.

Anderson, J. R., & Bower, G. H. (2014). *Human Associative Memory*. Psychology Press.

Andrä, C., Mathias, B., Schwager, A., Macedonia, M., & von Kriegstein, K. (2020). Learning Foreign Language Vocabulary with Gestures and Pictures Enhances Vocabulary Memory for Several Months Post-Learning in Eight-Year-Old School Children. *Educational Psychology Review*, *32*(3), 815–850. https://doi.org/10.1007/s10648-020-09527-z

Asher, J. J. (1969). The Total Physical Response Approach to Second Language Learning. *The Modern Language Journal*, *53*(1), 3–17. https://doi.org/10.2307/322091

Bahr, R. H., Silliman, E. R., & Berninger, V. W. (2020). Derivational Morphology Bridges Phonology and Orthography: Insights Into the Development of Word-Specific Spellings by Superior, Average, and Poor Spellers. *Language, Speech, and Hearing Services in Schools*, *51*(3), 640–654. https://doi.org/10.1044/2020_LSHSS-19-00090

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2015). Tracking lexical consolidation with ERPs: Lexical and semantic-priming effects on N400 and

LPC responses to newly-learned words. *Neuropsychologia*, *79*, 33–41. https://doi.org/10.1016/j.neuropsychologia.2015.10.020

Bara, B. G., & Tirassa, M. (1999). *A mentalist framework for linguistic and extralinguistic communication*.

Barbir, M., Babineau, M. J., Fiévet, A.-C., & Christophe, A. (2023). Rapid infant learning of syntactic–semantic links. *Proceedings of the National Academy of Sciences*, *120*(1), e2209153119. https://doi.org/10.1073/pnas.2209153119

Barcroft, J. (2007). Effects of Opportunities for Word Retrieval During Second Language Vocabulary Learning. *Language Learning*, *57*(1), 35–56. https://doi.org/10.1111/j.1467-9922.2007.00398.x

Beattie, G., & Shovelton, H. (1999). *Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation*. *123*(1–2), 1–30. https://doi.org/10.1515/semi.1999.123.1-2.1

Bergström, K., Klatte, M., Steinbrink, C., & Lachmann, T. (2016). First and Second Language Acquisition in German Children Attending a Kindergarten Immersion Program: A Combined Longitudinal and Cross-Sectional Study. *Language Learning*, *66*(2), 386–418. https://doi.org/10.1111/lang.12162

Billot-Vasquez, K., Lian, Z., Hirata, Y., & Kelly, S. (2020). Emblem Gestures Improve Perception and Evaluation of Non-native Speech. *Frontiers in Psychology*, *11*. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2020.574418

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service? *Bilingualism: Language and Cognition*, *13*(3), 359–371. https://doi.org/10.1017/S1366728909990344

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R. J., Zilles, K., Rizzolatti, G., & Freund, H.-J. (2004). Action Observation Activates Premotor and Parietal Areas in a Somatotopic Manner: An fMRI Study. In *Social Neuroscience*. Psychology Press.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*(3), 149–210. https://doi.org/10.1007/BF01320076

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*(3), 332–361. https://doi.org/10.1037/0033-295X.97.3.332

Cohen, R. L. (1981). On the generality of some memory laws. *Scandinavian Journal of Psychology*, *22*(1), 267–281. https://doi.org/10.1111/j.1467-9450.1981.tb00402.x

Cohen, R. L. (1989). Memory for action events: The power of enactment. *Educational Psychology Review*, *1*(1), 57–80. https://doi.org/10.1007/BF01326550

Comesaña, M., Perea, M., Piñeiro, A., & Fraga, I. (2009). Vocabulary teaching strategies and conceptual representations of words in L2 in children: Evidence with novice learners. *Journal of Experimental Child Psychology*, *104*(1), 22–33. https://doi.org/10.1016/j.jecp.2008.10.004

Congdon, E. L., Novack, M. A., Brooks, N., Hemani-Lopez, N., O'Keefe, L., & Goldin-Meadow, S. (2017). Better together: Simultaneous presentation of speech and gesture in math instruction supports generalization and retention. *Learning and Instruction*, *50*, 65–74. https://doi.org/10.1016/j.learninstruc.2017.03.005

Conrad, C. (1974). Context effects in sentence comprehension: A study of the subjective lexicon. *Memory & Cognition*, *2*(1), 130–138. https://doi.org/10.3758/BF03197504

Cook, S. W., Duffy, R. G., & Fenn, K. M. (2013). Consolidation and Transfer of Learning After Observing Hand Gesture. *Child Development*, *84*(6), 1863–1871. https://doi.org/10.1111/cdev.12097

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

Dargue, N., Sweller, N., & Jones, M. P. (2019). When our hands help us understand: A meta-analysis into the effects of gesture on comprehension. *Psychological Bulletin*, *145*(8), 765–784. https://doi.org/10.1037/bul0000202

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3773–3800. https://doi.org/10.1098/rstb.2009.0111

De Groot, A. M. B., Dannenburg, L., & Vanhell, J. G. (1994). Forward and Backward Word Translation by Bilinguals. *Journal of Memory and Language*, *33*(5), 600–629. https://doi.org/10.1006/jmla.1994.1029

Deacon, D., Dynowska, A., Ritter, W., & Grose-Fifer, J. (2004). Repetition and semantic priming of nonwords: Implications for theories of N400 and word recognition. *Psychophysiology*, *41*(1), 60–74. https://doi.org/10.1111/1469-8986.00120

Dijkstra, T., Heuven, W. J. B. V., & Grainger, J. (1998). *Simulating Cross-Language Competition with the Bilingual Interactive Activation Model* (3–4). *38*(3–4), Article 3–4. https://doi.org/10.5334/pb.933

Dijkstra, T., & Van Heuven, W. J. B. (1998). The BIA Model and Bilingual Word Recognition. In *Localist Connectionist Approaches To Human Cognition*. Psychology Press.

Dijkstra, T., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175–197. https://doi.org/10.1017/S1366728902003012

Dittinger, E., Barbaroux, M., D'Imperio, M., Jäncke, L., Elmer, S., & Besson, M. (2016). Professional Music Training and Novel Word Learning: From Faster Semantic Encoding to Longer-lasting Word Representations. *Journal of Cognitive Neuroscience*, *28*(10), 1584–1602. https://doi.org/10.1162/jocn_a_00997

Dittinger, E., Scherer, J., Jäncke, L., Besson, M., & Elmer, S. (2019). Testing the influence of musical expertise on novel word learning across the lifespan using a cross-sectional approach in children, young adults and older adults. *Brain and Language*, *198*, 104678. https://doi.org/10.1016/j.bandl.2019.104678

Dufour, R., & Kroll, J. F. (1995). Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory & Cognition*, *23*(2), 166–180. https://doi.org/10.3758/BF03197219

Dumay, N., & Gaskell, M. G. (2007). Sleep-Associated Changes in the Mental Representation of Spoken Words. *Psychological Science*, *18*(1), 35–39. https://doi.org/10.1111/j.1467-9280.2007.01845.x

Emmorey, K., Borinstein, H. B., Thompson, R., & Gollan, T. H. (2008). Bimodal bilingualism. *Bilingualism: Language and Cognition*, *11*(1), 43–61. https://doi.org/10.1017/S1366728907003203

Engelkamp, J. (1986). Motor Programs as Part of the Meaning of Verbal Items. In I. Kurcz, G. W. Shugar, & J. H. Danks (Eds.), *Advances in Psychology* (Vol. 39, pp. 115–138). North-Holland. https://doi.org/10.1016/S0166-4115(09)60135-7

Engelkamp, J. (2001). Action Memory: A System-Oriented Approach. In H. D. Zimmer, R. L. Cohen, M. J. Guynn, J. Engelkamp, R. Kormi-Nouri, & M. A. Foley (Eds.), *Memory for Action: A Distinct Form of Episodic Memory?* (p. 0). Oxford University Press. https://doi.org/10.1093/oso/9780195115536.003.0003

Engelkamp, J., & Dehn, D. M. (2000). Item and order information in subject-performed tasks and experimenter-performed tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 671–682. https://doi.org/10.1037/0278-7393.26.3.671

Engelkamp, J., & Zimmer, H. D. (1985). Motor programs and their relation to semantic memory. *German Journal of Psychology*, *9*(3), 239–254.

Engelkamp, J., & Zimmer, H. D. (1997). Sensory factors in memory for subject-performed tasks. *Acta Psychologica*, *96*(1), 43–60. https://doi.org/10.1016/S0001-6918(97)00005-X

Engelkamp, J., Zimmer, H. D., Mohr, G., & Sellen, O. (1994). Memory of self-performed tasks: Self-performing during recognition. *Memory & Cognition*, *22*(1), 34–39. https://doi.org/10.3758/BF03202759

Fadiga, L., & Craighero, L. (2003). New insights on sensorimotor integration: From hand action to speech perception. *Brain and Cognition*, *53*(3), 514–524. https://doi.org/10.1016/S0278-2626(03)00212-4

Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). CONTEXT OF LEARNING AND SECOND LANGUAGE FLUENCY IN FRENCH: Comparing Regular Classroom, Study Abroad, and Intensive Domestic Immersion Programs. *Studies in Second Language Acquisition*, *26*(2), 275–301. https://doi.org/10.1017/S0272263104262064

Frick-Horbury, D. (2002). The Effects of Hand Gestures on Verbal Recall as a Function of High- and Low-Verbal-Skill Levels. *The Journal of General Psychology*, *129*(2), 137–147. https://doi.org/10.1080/00221300209603134

Gais, S., Lucas, B., & Born, J. (2006). Sleep after learning aidsmemory recall. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *13*, 259–262. https://doi.org/10.1101/lm.132106

García-Gámez, A. B., Cervilla, Ó., Casado, A., & Macizo, P. (2021). Seeing or acting? The effect of performing gestures on foreign language vocabulary learning. *Language Teaching Research*, 13621688211024364. https://doi.org/10.1177/13621688211024364

García-Gámez, A. B., & Macizo, P. (2019). Learning nouns and verbs in a foreign language: The role of gestures. *Applied Psycholinguistics*, *40*(2), 473–507. https://doi.org/10.1017/S0142716418000656

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*(2), 105–132. https://doi.org/10.1016/S0010-0277(03)00070-2

Geukes, S., Gaskell, M. G., & Zwitserlood, P. (2015). Stroop effects from newly learned color words: Effects of memory consolidation and episodic context. *Frontiers in Psychology*, *6*. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00278

Goldstein, M. H., & Schwade, J. A. (2008). Social Feedback to Infants' Babbling Facilitates Rapid Phonological Learning. *Psychological Science*, *19*(5), 515–523. https://doi.org/10.1111/j.1467-9280.2008.02117.x

Gough, V. (2023). *The British and their secret language regrets | British Council*. https://www.britishcouncil.org/voices-magazine/british-and-their-secret-language-regrets

Guan, G., Hu, A., Guo, L., & Yu, H. (2019). *An Eye-tracking Technique—Visual-World Paradigm*. 455–457. https://doi.org/10.2991/icedem-19.2019.108

Gullberg, M. (2006). Handling Discourse: Gestures, Reference Tracking, and Communication Strategies in Early L2. *Language Learning*, *56*(1), 155–196. https://doi.org/10.1111/j.0023-8333.2006.00344.x

Gullberg, M. (2008). Gestures and second language acquisition. *Handbook of Cognitive Linguistics and Second Language Acquisition*, 276–305.

Gullberg, M., Roberts, L., & Dimroth, C. (2012). What word-level knowledge can adult learners acquire after minimal exposure to a new language? *IRAL, International Review of Applied Linguistics in Language Teaching*, *50*(4), 239–276.

Halvorson, K. M., Bushinski, A., & Hilverman, C. (2019). The role of motor context in the beneficial effects of hand gesture on memory. *Attention, Perception, & Psychophysics*, *81*(7), 2354–2364. https://doi.org/10.3758/s13414-019-01734-3

Hartsuiker, R. J., & Moors, A. (2017). On the automaticity of language processing. In *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge* (pp. 201–225). De Gruyter Mouton. https://doi.org/10.1037/15969-010

Hayes-Harb, R., & Barrios, S. (2021). The influence of orthography in second language phonological acquisition. *Language Teaching*, *54*(3), 297–326. https://doi.org/10.1017/S0261444820000658

Helstrup, T. (1987). One, two, or three memories? A problem-solving approach to memory for performed acts. *Acta Psychologica*, *66*(1), 37–68. https://doi.org/10.1016/0001-6918(87)90017-5

Hinojosa, J. A., Martín-Loeches, M., & Rubia, F. J. (2001). Event-Related Potentials and Semantics: An Overview and an Integrative Proposal. *Brain and Language*, *78*(1), 128–139. https://doi.org/10.1006/brln.2001.2455

Holle, H., & Gunter, T. C. (2007). The Role of Iconic Gestures in Speech Disambiguation: ERP Evidence. *Journal of Cognitive Neuroscience*, *19*(7), 1175–1192. https://doi.org/10.1162/jocn.2007.19.7.1175

Holler, J., & Beattie, G. (2003). Pragmatic aspects of representational gestures: Do speakers use them to clarify verbal ambiguity for the listener? *Gesture*, *3*(2), 127–154. https://doi.org/10.1075/gest.3.2.02hol

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. https://doi.org/10.1016/j.cognition.2007.02.001

Hostetter, A. (2011). When Do Gestures Communicate? A Meta-Analysis. *Psychological Bulletin*, *137*, 297–315. https://doi.org/10.1037/a0022128

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495–514. https://doi.org/10.3758/PBR.15.3.495

Huang, X., Kim, N., & Christianson, K. (2019). Gesture and Vocabulary Learning in a Second Language. *Language Learning*, *69*(1), 177–197. https://doi.org/10.1111/lang.12326

Hussain, Z., Sekuler, A. B., & Bennett, P. J. (2009). How much practice is needed to produce perceptual learning? *Vision Research*, *49*(21), 2624–2634. https://doi.org/10.1016/j.visres.2009.08.022

Ianì, F., & Bucciarelli, M. (2017). Mechanisms underlying the beneficial effect of a speaker's gestures on the listener. *Journal of Memory and Language*, *96*, 110–121. https://doi.org/10.1016/j.jml.2017.05.004

Jared, D., Poh, R. P. Y., & Paivio, A. (2013). L1 and L2 picture naming in Mandarin–English bilinguals: A test of Bilingual Dual Coding Theory. *Bilingualism: Language and Cognition*, *16*(2), 383–396. https://doi.org/10.1017/S1366728912000685

Jiang, N. (2015). Six decades of research on lexical representationand processing in bilinguals. In J. W. Schwieter (Ed.), *The Cambridge Handbook of Bilingual Processing* (pp. 29–84). Cambridge University Press. https://doi.org/10.1017/CBO9781107447257.002

Kandana Arachchige, K. G., Holle, H., Rossignol, M., Loureiro, I. S., & Lefebvre, L. (2022). High verbal working memory load impairs gesture-speech integration: Evidence from a dual task paradigm. *Gesture*, *20*(3), 354–375. https://doi.org/10.1075/gest.20028.kan

Kandana Arachchige, K. G., Simoes Loureiro, I., Blekic, W., Rossignol, M., & Lefebvre, L. (2021). The Role of Iconic Gestures in Speech Comprehension: An Overview of Various Methodologies. *Frontiers in Psychology*, *12*. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.634074

Kapnoula, E. C., Packard, S., Gupta, P., & McMurray, B. (2015). Immediate lexical integration of novel word forms. *Cognition*, *134*, 85–99. https://doi.org/10.1016/j.cognition.2014.09.007

Kassim, W. Z. W. (2018). Utilizing Dual Coding Theory and animated images to enhance ESL students' vocabulary learning. *The English Teacher*, *47*(3), Article 3. https://meltajournals.com/index.php/TET/article/view/48

Kelly, S., Creigh, P., & Bartolotti, J. (2010). Integrating Speech and Iconic Gestures in a Stroop-like Task: Evidence for Automatic Processing. *Journal of Cognitive Neuroscience*, *22*(4), 683–694. Journal of Cognitive Neuroscience. https://doi.org/10.1162/jocn.2009.21254

Kelly, S., Healey, M., Özyürek, A., & Holler, J. (2015). The processing of speech, gesture, and action during language comprehension. *Psychonomic Bulletin & Review*, *22*(2), 517–523. https://doi.org/10.3758/s13423-014-0681-7

Kelly, S., & Lee, A. (2012). When actions speak too much louder than words: Hand gestures disrupt word learning when phonetic demands are high. *Language and Cognitive Processes*, *27*(6), 793–807. https://doi.org/10.1080/01690965.2011.581125

Kelly, S., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, *24*(2), 313–334. https://doi.org/10.1080/01690960802365567

Kelly, S., Özyürek, A., & Maris, E. (2010). Two Sides of the Same Coin: Speech and Gesture Mutually Interact to Enhance Comprehension. *Psychological Science*, *21*(2), 260–267. https://doi.org/10.1177/0956797609357327

Kelly, S., Ward, S., Creigh, P., & Bartolotti, J. (2007). An intentional stance modulates the integration of gesture and speech during comprehension. *Brain and Language*, *101*(3), 222–233. https://doi.org/10.1016/j.bandl.2006.07.008

Kotz, S. A., & Elston-Güttler, K. (2004). The role of proficiency on processing categorical and associative information in the L2 as revealed by reaction times and event-related brain potentials. *Journal of Neurolinguistics*, *17*(2), 215–235. https://doi.org/10.1016/S0911-6044(03)00058-7

Kroll, J. F., Hell, J. G. V., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition*, *13*(3), 373–381. https://doi.org/10.1017/S136672891000009X

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. https://doi.org/10.1006/jmla.1994.1008

Kroll, J. F., & Tokowicz, N. (2005). Models of Bilingual Representation and Processing: Looking Back and to the Future. In J. F. Kroll & A. M. B. D. Groot (Eds.), *Handbook of Bilingualism* (pp. 531–554). Oxford University PressNew York, NY. https://doi.org/10.1093/oso/9780195151770.003.0030

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science (New York, N.Y.)*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. MIT Press. https://doi.org/10.7551/mitpress/6393.001.0001

Lin, Y.-L. (2021). Gestures as scaffolding for L2 narrative recall: The role of gesture type, task complexity, and working memory. *Language Teaching Research*, 13621688211044584. https://doi.org/10.1177/13621688211044584

Lindsay, S., & Gaskell, M. G. (2010). A Complementary Systems Account of Word Learning in L1 and L2. *Language Learning*, *60*(s2), 45–63. https://doi.org/10.1111/j.1467-9922.2010.00600.x

Lindsay, S., & Gaskell, M. G. (2013). Lexical integration of novel words without sleep. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *39*(2), 608–622. https://doi.org/10.1037/a0029243

Lockhart, R. S., & Craik, F. I. M. (1990). Levels of processing: A retrospective commentary on a framework for memory research: Canadian Journal of Psychology / Revue canadienne de psychologie. *Canadian Journal of Psychology / Revue Canadienne de Psychologie*, *44*(1), 87–112. https://doi.org/10.1037/h0084237

Long, R., & Danechi, S. (2019). *English Baccalaureate*. https://commonslibrary.parliament.uk/research-briefings/sn06045/

Macedonia, M. (2014). Bringing back the body into the mind: Gestures enhance word learning in foreign language. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01467

Macedonia, M., & Klimesch, W. (2014). Long-Term Effects of Gestures on Memory for Foreign Language Words Trained in the Classroom. *Mind, Brain, and Education*, *8*(2), 74–88. https://doi.org/10.1111/mbe.12047

Macedonia, M., & Mueller, K. (2016). Exploring the Neural Representation of Novel Words Learned through Enactment in a Word Recognition Task. *Frontiers in Psychology*, *7*. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2016.00953

Macedonia, M., Müller, K., & Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping*, *32*(6), 982–998. https://doi.org/10.1002/hbm.21084

Macedonia, M., & Repetto, C. (2017). Why Your Body Can Jog Your Mind. *Frontiers in Psychology*, *8*, 362. https://doi.org/10.3389/fpsyg.2017.00362

Macedonia, M., & Von Kriegstein, K. (2012). Gestures Enhance Foreign Language Learning. *Biolinguistics*, *6*(3–4), 393–416. https://doi.org/10.5964/bioling.8931

Madan, C. R., & Singhal, A. (2012). Using actions to enhance memory: Effects of enactment, gestures, and exercise on human memory. *Frontiers in Psychology*, *3*. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2012.00507

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, *132*(2), 202–227. https://doi.org/10.1037/0096-3445.132.2.202

Mavilidi, M.-F., Okely, A. D., Chandler, P., Cliff, D. P., & Paas, F. (2015). Effects of Integrated Physical Exercises and Gestures on Preschool Children's Foreign Language Vocabulary Learning. *Educational Psychology Review*, *27*(3), 413–426. https://doi.org/10.1007/s10648-015-9337-z

Mayer, K. M., Yildiz, I. B., Macedonia, M., & von Kriegstein, K. (2015). Visual and Motor Cortices Differentially Support the Translation of Foreign Language Words. *Current Biology*, *25*(4), 530–535. https://doi.org/10.1016/j.cub.2014.11.068

McCafferty, S. G., & Stam, G. (2009). *Gesture: Second Language Acquistion and Classroom Research*. Routledge.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. https://doi.org/10.1037/0033-295X.102.3.419

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*(5), 375–407. https://doi.org/10.1037/0033-295X.88.5.375

McNamara, T. P. (2005). *Semantic Priming: Perspectives from Memory and Word Recognition*. Psychology Press. https://doi.org/10.4324/9780203338001

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, *92*(3), 350–371. https://doi.org/10.1037/0033-295X.92.3.350

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought* (pp. xi, 416). University of Chicago Press.

Nicolas, S., Ehrlich, M.-F., & Facci, G. (1996). Implicit memory and aging: Generation effect in a word-stem completion test. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, *15*(5), 513–533.

Nilsson, L. G., Nyberg, L., Klingberg, T., Aberg, C., Persson, J., & Roland, P. E. (2000). Activity in motor areas while remembering action events. *Neuroreport*, *11*(10), 2199–2201. https://doi.org/10.1097/00001756-200007140-00027

Nyberg, L., Petersson, K. M., Nilsson, L.-G., Sandblom, J., Åberg, C., & Ingvar, M. (2001). Reactivation of Motor Brain Areas during Explicit Memory for Actions. *NeuroImage*, *14*(2), 521–528. https://doi.org/10.1006/nimg.2001.0801

Obermeier, C., Dolk, T., & Gunter, T. C. (2012). The benefit of gestures during communication: Evidence from hearing and hearing-impaired individuals. *Cortex*, *48*(7), 857–870. https://doi.org/10.1016/j.cortex.2011.02.007

O'Reilly, R. C., Bhattacharyya, R., Howard, M. D., & Ketz, N. (2014). Complementary learning systems. *Cognitive Science*, *38*(6), 1229–1248. https://doi.org/10.1111/j.1551-6709.2011.01214.x

Ortega, G., & Özyürek, A. (2020). Types of iconicity and combinatorial strategies distinguish semantic categories in silent gesture across cultures. *Language and Cognition*, *12*(1), 84–113. https://doi.org/10.1017/langcog.2019.28

Paivio, A. (1978). A Dual Coding Approach to Perception and Cognition. In *Modes of Perceiving and Processing Information*. Psychology Press.

Paivio, A. (1990). *Mental Representations: A dual coding approach*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195066661.001.0001

Paivio, A., & Csapo, K. (1969). Concrete image and verbal memory codes. *Journal of Experimental Psychology*, *80*(2, Pt.1), 279–285. https://doi.org/10.1037/h0027273

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology*, *5*(2), 176–206. https://doi.org/10.1016/0010-0285(73)90032-7

Paivio, A., & Desrochers, A. (1980). A dual-coding approach to bilingual memory. *Canadian Journal of Psychology*, *34*, 388–399.

Palma, P., & Titone, D. (2021). Something old, something new: A review of the literature on sleep-related lexicalization of novel words in adults. *Psychonomic Bulletin & Review*, *28*(1), 96–121. https://doi.org/10.3758/s13423-020-01809-5

Palmer, S. D., & Havelka, J. (2010). Age of acquisition effects in vocabulary learning. *Acta Psychologica*, *135*(3), 310–315. https://doi.org/10.1016/j.actpsy.2010.08.002

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Polis. (2024). *The Polis Method.* The Jerusalem Institute of Languages and Humanities. https://www.polisjerusalem.org/about/the-polis-method/

Porter, A. (2012). A helping hand with language learning: Teaching French vocabulary with gesture. *The Language Learning Journal*, *44*(2), 236–256. https://doi.org/10.1080/09571736.2012.750681

Potter, M. C., So, K.-F., Eckardt, B. V., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior*, *23*(1), 23–38. https://doi.org/10.1016/S0022-5371(84)90489-4

Prince, P. (1996). Second Language Vocabulary Learning: The Role of Context versus Translations as a Function of Proficiency. *The Modern Language Journal*, *80*(4), 478–493. https://doi.org/10.2307/329727

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510–532. https://doi.org/10.1037/0033-2909.114.3.510

Repetto, C., Mathias, B., Weichselbaum, O., & Macedonia, M. (2021). Visual recognition of words learned with gestures induces motor resonance in the forearm muscles. *Scientific Reports*, *11*(1), 17278. https://doi.org/10.1038/s41598-021-96792-9

Repetto, C., Pedroli, E., & Macedonia, M. (2017). Enrichment Effects of Gestures and Pictures on Abstract Words in a Second Language. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/article/10.3389/fpsyg.2017.02136

Rizzolatti, G., & Craighero, L. (2004). THE MIRROR-NEURON SYSTEM. *Annual Review of Neuroscience*, *27*(1), 169–192. https://doi.org/10.1146/annurev.neuro.27.070203.144230

Rogers, W. T. (1978). The Contribution of Kinesic Illustrators toward the Comprehension of Verbal Behavior within Utterances. *Human Communication Research*, *5*(1), 54–62. https://doi.org/10.1111/j.1468-2958.1978.tb00622.x

Sadeghi, K. (2009). Collocational Differences Between L1 and L2: Implications for EFL Learners and Teachers. *TESL Canada Journal*, 100–124. https://doi.org/10.18806/tesl.v26i2.417

Samuels, S. J., & Flor, R. F. (1997). The Importance of Automaticity for Developing Expertise in Reading. *Reading & Writing Quarterly*, *13*(2), 107–121. https://doi.org/10.1080/1057356970130202

Schimke, E. A. E., Angwin, A. J., Cheng, B. B. Y., & Copland, D. A. (2021). The effect of sleep on novel word learning in healthy adults: A systematic review and meta-analysis. *Psychonomic Bulletin & Review*, *28*(6), 1811–1838. https://doi.org/10.3758/s13423-021-01980-3

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66. https://doi.org/10.1037/0033-295X.84.1.1

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–190. https://doi.org/10.1037/0033-295X.84.2.127

Silverberg, S., & Samuel, A. G. (2004). The effect of age of second language acquisition on the representation and processing of second language words. *Journal of Memory and Language*, *51*(3), 381–398. https://doi.org/10.1016/j.jml.2004.05.003

Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science*, *16*(2), 85–89. https://doi.org/10.1111/j.0956-7976.2005.00786.x

Sivashankar, Y., Liu, J., & Fernandes, M. A. (2023). The importance of performing versus observing meaningful actions, on the enactment benefit to memory. *Journal of Cognitive Psychology*, *35*(1), 47–58. https://doi.org/10.1080/20445911.2022.2102639

Smotrova, T., & Lantolf, J. P. (2013). The Function of Gesture in Lexically Focused L2 Instructional Conversations. *The Modern Language Journal*, *97*(2), 397–416.

So, W. C., Sim Chen-Hui, C., & Low Wei-Shan, J. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall? *Language and Cognitive Processes*, *27*(5), 665–681. https://doi.org/10.1080/01690965.2011.573220

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Sueyoshi, A., & Hardison, D. M. (2005). The Role of Gestures and Facial Cues in Second Language Listening Comprehension. *Language Learning*, *55*(4), 661–699. https://doi.org/10.1111/j.0023-8333.2005.00320.x

Sunderman, G., & Kroll, J. F. (2006). FIRST LANGUAGE ACTIVATION DURING SECOND LANGUAGE LEXICAL PROCESSING: An Investigation of Lexical Form, Meaning, and Grammatical Class. *Studies in Second Language Acquisition*, *28*(3), 387–422. https://doi.org/10.1017/S0272263106060177

Talamas, A., Kroll, J. F., & Dufour, R. (1999). From form to meaning: Stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition*, *2*(1), 45–58. https://doi.org/10.1017/S1366728999000140

Tamminen, J., & Gaskell, M. G. (2008). Short Article: Newly Learned Spoken Words Show Long-Term Lexical Competition Effects. *Quarterly Journal of Experimental Psychology*, *61*(3), 361–371. https://doi.org/10.1080/17470210701634545

Tamminen, J., & Gaskell, M. G. (2013). Novel word integration in the mental lexicon: Evidence from unmasked and masked semantic priming. *Quarterly Journal of Experimental Psychology*, *66*(5), 1001–1025. https://doi.org/10.1080/17470218.2012.724694

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep Spindle Activity is Associated with the Integration of New Memories and Existing Knowledge. *Journal of Neuroscience*, *30*(43), 14356–14360. https://doi.org/10.1523/JNEUROSCI.3028-10.2010

Tellier, M. (2007). How do teacher's gestures help young children in second language acquisition? *International Society of Gesture Studies, ISGS*, en ligne. https://hal.science/hal-00433380

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, *8*(2), 219–235. https://doi.org/10.1075/gest.8.2.06tel

Tham, E. K. H., Lindsay, S., & Gaskell, M. G. (2015). Markers of automaticity in sleep-associated consolidation of novel words. *Neuropsychologia*, *71*, 146–157. https://doi.org/10.1016/j.neuropsychologia.2015.03.025

Van Der Ven, F., Takashima, A., Segers, E., & Verhoeven, L. (2015). Learning Word Meanings: Overnight Integration and Study Modality Effects. *PLOS ONE*, *10*(5), e0124926. https://doi.org/10.1371/journal.pone.0124926

Véliz-Campos, M., Villegas, C., & Veliz, L. (2023). Foreign Language Anxiety in Online and In-Person Learning Environments: A Case from Chile. *Language Teaching Research Quarterly*, *35*, 37–56.

Walker, S., Henderson, L. M., Fletcher, F. E., Knowland, V. C. P., Cairney, S. A., & Gaskell, M. G. (2019). Learning to live with interfering neighbours: The influence of time of learning and level of encoding on word learning. *Royal Society Open Science*, *6*(4), 181842. https://doi.org/10.1098/rsos.181842

Wang, M., Perfetti, C. A., & Liu, Y. (2003). Alphabetic Readers Quickly Acquire Orthographic Structure in Learning to Read Chinese. *Scientific Studies of Reading*, *7*(2), 183–208. https://doi.org/10.1207/S1532799XSSR0702_4

Warren, R. E. (1972). Stimulus encoding and memory. *Journal of Experimental Psychology*, *94*(1), 90–100. https://doi.org/10.1037/h0032786

Warren, R. E. (1974). Association, directionality, and stimulus encoding. *Journal of Experimental Psychology*, *102*(1), 151–158. https://doi.org/10.1037/h0035703

Weighall, A. R., Henderson, L. M., Barr, D. J., Cairney, S. A., & Gaskell, M. G. (2017). Eye-tracking the time-course of novel word learning and lexical competition in adults and children. *Brain and Language*, *167*, 13–27. https://doi.org/10.1016/j.bandl.2016.07.010

Whelan, R. (2008). Effective Analysis of Reaction Time Data. *The Psychological Record*, *58*(3), 475–482. https://doi.org/10.1007/BF03395630

Willems, R. M., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain and Language*, *101*(3), 278–289. https://doi.org/10.1016/j.bandl.2007.03.004

Zhao, W., Li, Y., & Du, Y. (2021). TMS Reveals Dynamic Interaction between Inferior Frontal Gyrus and Posterior Middle Temporal Gyrus in Gesture-Speech Semantic

Integration. *Journal of Neuroscience*, *41*(50), 10356–10364. https://doi.org/10.1523/JNEUROSCI.1355-21.2021

Zhao, W., Riggs, K., Schindler, I., & Holle, H. (2018). Transcranial Magnetic Stimulation over Left Inferior Frontal and Posterior Temporal Cortex Disrupts Gesture-Speech Integration. *The Journal of Neuroscience*, *38*(8), 1891–1900. https://doi.org/10.1523/JNEUROSCI.1748-17.2017

Zimmer, H. D. (2001). *Memory for action: A distinct form of episodic memory?.* Oxford University Press, USA.

APPENDIX A

Pretest studies: mean reaction times for native and non-native speakers

	Mean RT native (SD)		Mean RT non-native (SD)	
	Semantically Congruent	Semantically Incongruent	Semantically Congruent	Semantically Incongruent
Gender Congruent	723 (233)	731 (235)	621 (202)	631 (204)
Gender Incongruent	726 (238)	738 (235)	628 (202)	638 (200)

Note. Time in milliseconds (ms)

APPENDIX B

Mandarin and English vocabulary word list, blocks and pairs

Table B1

Vocabulary list

Mandarin word	English translation
guān	close
jù	saw
să	sprinkle
diăn rán	light
găn	roll
chuí	hammer
jiăn	cut
dă zì	type
jié	knot
mŏ	spread
dào	pour
kāi suŏ	unlock
pēn	spray
kāi	open
jiăo bàn	whisk
săo	sweep
rēng	throw
jĭ	squeeze
nòng duàn	snap
xiĕ	write
sī	tear
bō	dial
yáo	shake
lā shàng	zip up

Table B2

Vocabulary learning blocks

List A		List B	
Block 1		**Block 4**	
Close	guān	Spray	pēn
Saw	jù	Open	kāi
Sprinkle	sǎ	Whisk	jiǎo bàn
Light	diǎn rán	Sweep	sǎo
Block 2		**Block 5**	
Roll	gǎn	Throw	rēng
Hammer	chuí	Squeeze	jǐ
Cut	jiǎn	Snap	nòng duàn
Type	dǎ zì	Write	xiě
Block 3		**Block 6**	
Knot	jié	Tear	sī
Spread	mǒ	Dial	bō
Pour	dào	Shake	yáo
Unlock	kāi suǒ	Zip up	lā shàng

Table B3

Word pairings for the action-speech integration task

Pairs	English words	Mandarin translation
1	hammer & close	chuí & guān
2	type & saw	dǎ zì & jù
3	sprinkle & knot	sǎ & jié
4	light & spread	diǎn rán & mǒ
5	cut & unlock	jiǎn & kāi suǒ
6	pour & roll	dào & gǎn
7	spray and squeeze	pēn & jǐ
8	throw & open	rēng & kāi
9	whisk & tear	jiǎo bàn & sī
10	sweep & dial	sǎo & bō
11	snap & shake	nòng duàn & yáo
12	zip up & write	lā shàng & xiě

APPENDIX C

Semantic congruency effects for word pairs (native pre-test study)

Pair	Actions	Semantic congruency effect (ms)
2	Hammer & Spray	28.34
15	Light & Turn	28.01
1	Wash & Knock	25.20
21	Scrub & Roll	22.56
11	Steer & Zip up	20.24
13	Brush & Write	19.24
3	Unlock & Sprinkle	15.96
14	Chop & Squeeze	13.59
16	Pour & Dial	13.00
10	Paint & Snap	12.62
12	Crack & Sweep	9.54
18	Saw & Wipe	9.24
6	Knot & Unwrap	8.37
8	Whisk & Throw	7.91
9	Grate & Type	6.12
17	Cut & Shake	5.91
22	Unscrew & Spread	3.68
19	Pull & Stir	1.38
7	Peel & Click	1.15
4	Tear & Close	-1.23
5	Sharpen & Flip	-4.75
20	Wring out & Open	-6.18

APPENDIX D

Mini test results for Studies 1-6

Table D1

Study 1 mini test results

Participant	Mini test raw score/96	Mini test score %
1	94	97.92
2	56	58.33
3	89	92.71
4	94	97.92
5	90	93.75
6	95	98.96
7	91	94.79
8	92	95.83
9	96	100.00
10	94	97.92
11	91	94.79
12	95	98.96
13	96	100.00
14	76	79.17
15	92	95.83
16	92	95.83
17	91	94.79
18	94	97.92
19	96	100.00
20	96	100.00
21	96	100.00
22	79	82.29
23	95	98.96
24	90	93.75
25	81	84.38
26	82	85.42
27	91	94.79
28	43	44.79
29	96	100.00
30	96	100.00
31	96	100.00
32	96	100.00
33	78	81.25
34	38	39.58
35	96	100.00

36	83	86.46
37	95	98.96
38	54	56.25
39	96	100.00
40	96	100.00
41	96	100.00
42	95	98.96
43	92	95.83
44	95	98.96
45	95	98.96
46	95	98.96
47	95	98.96
48	82	85.42
49	90	93.75
50	96	100.00
51	71	73.96
52	69	71.88
53	92	95.83
54	94	97.92

Note. Mean score = 87.85, 2 SD below the mean = 61.12

Table D2

Study 2 mini test results

Participant	Mini test raw score/96	Mini test score %
1	95	98.96
2	94	97.92
3	93	96.88
4	93	96.88
5	87	90.63
6	96	100.00
7	92	95.83
8	89	92.71
9	91	94.79
10	73	76.04
11	90	93.75
12	84	87.50
13	92	95.83
14	94	97.92
15	78	81.25
16	88	91.67
17	95	98.96

18	95	98.96
19	78	81.25
20	94	97.92
21	95	98.96
22	94	97.92
23	83	86.46
24	96	100.00
25	66	68.75
26	95	98.96
27	95	98.96
28	78	81.25
29	94	97.92
30	84	87.50
31	88	91.67
32	90	93.75
33	68	70.83
34	81	84.38
35	95	98.96
36	87	90.63
37	67	69.79
38	90	93.75
39	77	80.21
40	90	93.75
41	94	97.92
42	69	71.88
43	90	93.75
44	93	96.88
45	94	97.92
46	94	97.92
47	93	96.88
48	91	94.79
49	91	94.79
50	92	95.83
51	96	100.00
52	80	83.33
53	95	98.96
54	93	96.88
55	91	94.79
56	94	97.92
57	92	95.83
58	86	89.58
59	91	94.79
60	86	89.58
61	88	91.67
62	87	90.63

63	81	84.38
64	81	84.38
65	93	96.88
66	83	86.46
67	86	89.58
68	94	97.92
69	93	96.88
70	77	80.21
71	94	97.92
72	87	90.63
73	95	98.96
74	94	97.92
75	96	100.00
76	94	97.92

Note. Mean score = 88.58, 2 SD below the mean = 73.60

Table D3

Study 3 mini test results

Participant	Mini test raw score/96	Mini test score %
1	95	98.96
2	92	95.83
3	89	92.71
4	93	96.88
5	94	97.92
6	92	95.83
7	83	86.46
8	91	94.79
9	92	95.83
10	74	77.08
11	70	72.92
12	93	96.88
13	91	94.79
14	83	86.46
15	95	98.96
16	88	91.67
17	85	88.54
18	94	97.92
19	80	83.33
20	87	90.63

21	94	97.92
22	92	95.83
23	89	92.71
24	80	83.33
25	84	87.50
26	96	100.00
27	87	90.63
28	95	98.96
29	96	100.00
30	93	96.88
31	94	97.92
32	95	98.96
33	76	79.17
34	88	91.67
35	92	95.83
36	94	97.92
37	95	98.96
38	92	95.83
39	95	98.96
40	91	94.79
41	95	98.96
42	96	100.00
43	85	88.54
44	84	87.50
45	88	91.67
46	90	93.75
47	86	89.58
48	80	83.33

Note. Mean score = 89.23, 2 SD below the mean = 76.82

Table D4

Study 4 mini test results

Participant	Mini test raw score/96	Mini test score %
1	96	100.00
2	87	90.63
3	94	97.92
4	93	96.88
5	94	97.92
6	94	97.92

7	80	83.33
8	91	94.79
9	96	100.00
10	95	98.96
11	95	98.96
12	91	94.79
13	96	100.00
14	89	92.71
15	88	91.67
16	94	97.92
17	86	89.58
18	92	95.83
19	86	89.58
20	91	94.79
21	96	100.00
22	94	97.92
23	92	95.83
24	93	96.88
25	91	94.79
26	93	96.88
27	90	93.75
28	95	98.96
29	95	98.96
30	83	86.46
31	94	97.92
32	94	97.92
33	94	97.92
34	88	91.67
35	89	92.71
36	79	82.29
37	96	100.00
38	89	92.71
39	91	94.79
40	95	98.96
41	91	94.79
42	96	100.00
43	84	87.50
44	92	95.83
45	89	92.71
46	94	97.92

Note. Mean score = 91.41, 2 SD below the mean = 83.00

Table D5

Study 5 mini test results

Participant	Mini test raw score/48	Mini test score %
1	47	97.92
2	48	100.00
3	47	97.92
4	48	100.00
5	48	100.00
6	47	97.92
7	48	100.00
8	25	52.08
9	48	100.00
10	48	100.00
11	48	100.00
12	45	93.75
13	47	97.92
14	43	89.58
15	43	89.58
16	48	100.00
17	45	93.75
18	48	100.00
19	48	100.00
20	48	100.00
21	48	100.00
22	45	93.75
23	45	93.75
24	45	93.75
25	47	97.92
26	47	97.92
27	48	100.00
28	44	91.67
29	48	100.00
30	40	83.33
31	48	100.00
32	47	97.92
33	36	75.00
34	48	100.00
35	43	89.58
36	48	100.00
37	44	91.67
38	46	95.83
39	45	93.75

40	40	83.33
41	48	100.00
42	46	95.83
43	48	100.00
44	48	100.00
45	43	89.58
46	48	100.00
47	43	89.58
48	44	91.67
49	48	100.00
50	44	91.67
51	48	100.00
52	42	87.50
53	48	100.00
54	48	100.00

Note. Mean score = 45.65, 2 SD below the mean = 37.90

Table D6

Study 6 mini test results

Participant	Mini test raw score /96	Mini test score %
1	93	96.88
2	95	98.96
3	96	100.00
4	86	89.58
5	96	100.00
6	90	93.75
7	96	100.00
8	95	98.96
9	94	97.92
10	90	93.75
11	91	94.79
12	78	81.25
13	94	97.92
14	89	92.71
15	92	95.83
16	93	96.88
17	96	100.00
18	89	92.71
19	96	100.00

20	95	98.96
21	93	96.88
22	94	97.92
23	96	100.00
24	84	87.50
25	94	97.92
26	93	96.88
27	93	96.88
28	96	100.00
29	86	89.58
30	73	76.04
31	94	97.92
32	96	100.00
33	96	100.00
34	88	91.67
35	74	77.08
36	94	97.92
37	94	97.92
38	96	100.00
39	96	100.00
40	92	95.83
41	90	93.75
42	94	97.92
43	82	85.42
44	95	98.96
45	94	97.92
46	89	92.71
47	75	78.13
48	94	97.92
49	69	71.88

Note. Mean score = 90.78, 2 SD below the mean = 77.31

APPENDIX E

Example data reviewing the learning effects for the word pairs for Study 1 gesture condition

Semantic congruency effect for word pairs across (Study 1, day 1, gesture condition only)

Pair	Words	Semantic congruency effect (ms)
4	Light & Spread	47
9	Tear & Whisk	29
11	Shake & Snap	20
12	Write & Zip up	17
3	Knot & Sprinkle	15
8	Throw & Open	14
10	Sweep & Dial	9
2	Saw & Type	4
6	Pour & Roll	4
7	Spray & Squeeze	4
5	Unlock & Cut	-18
1	Hammer & Close	-21

Note. The semantic congruency effect was calculated for each pair by taking the mean reaction time for the semantically incongruent stimuli for that pair and subtracting the mean reaction time for the semantically congruent conditions of that pair to find the difference.

Translation accuracy for word pairs (Study 1, gesture condition only)

Pair	Words	Day	Mean translation accuracy
12	Write & Zip up	1	0.80952381
8	Throw & Open	1	0.785714286
7	Spray & Squeeze	1	0.642857143
6	Pour & Roll	1	0.636363636
9	Tear & Whisk	1	0.619047619
11	Shake & Snap	1	0.571428571
4	Light & Spread	1	0.545454545
3	Knot & Sprinkle	1	0.522727273
5	Unlock & Cut	1	0.5
2	Saw & Type	1	0.454545455
10	Sweep & Dial	1	0.452380952
1	Hammer & Close	1	0.431818182

Note. Words paired according to the pairs used in the action-speech integration task.

APPENDIX F

Full table of results for the action-speech integration task – ANOVA for Studies 1-5

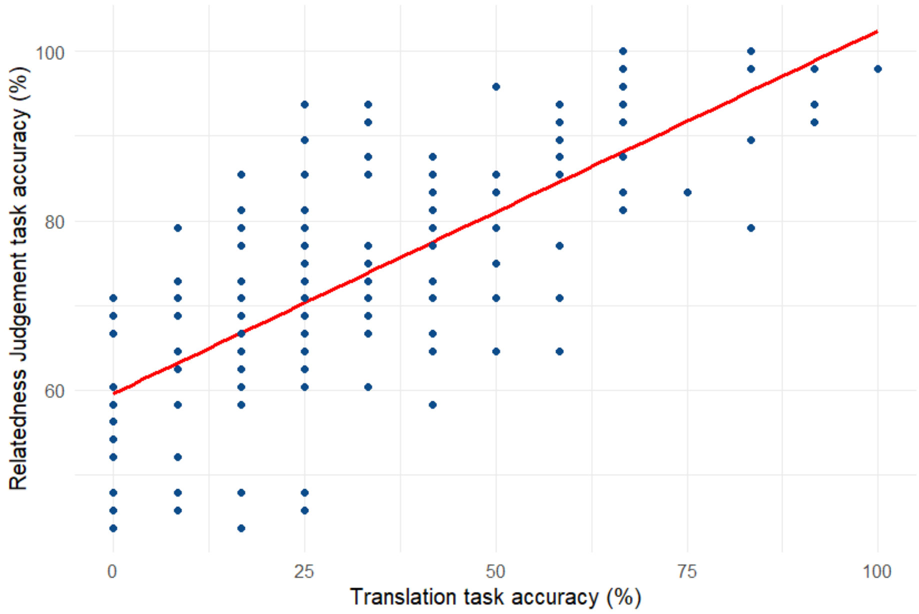
Effect	**df**	**MSE**	**F**	**p**
Study 1				
Learning method	1, 39	1701.59	0.14	0.71
Semantic congruency	1, 39	1638.96	0.32	0.575
Gender congruency	1, 39	1979.53	5.91	* .020
Day	1, 39	43326.61	2.34	0.134
Learning method:semantic congruency	1, 39	1484.44	4.76	* .035
Learning method:gender congruency	1, 39	1064.21	0.21	0.646
Semantic congruency:gender congruency	1, 39	949.67	0.03	0.861
Learning method:day	1, 39	1156.6	5.5	* .024
Semantic congruency:day	1, 39	801.13	0.56	0.46
Gender congruency:day	1, 39	1618.19	0.04	0.846
Learning method:semantic congruency:gender congruency	1, 39	1120.11	0.23	0.633
Learning method:semantic congruency:day	1, 39	1447.23	0.62	0.435
Learning method:gender congruency:day	1, 39	1344.93	3.48	0.07
Semantic congruency:gender congruency:day	1, 39	1616.64	0.12	0.734
Learning method:semantic congruency:gender congruency:day	1, 39	1345.3	0.33	0.568
Study 2				
Learning method	1, 39	1593.69	0.12	0.731
Semantic congruency	1, 39	1158.4	2.64	0.112
Gender congruency	1, 39	1357.07	18.27	* <.001
Day	1, 39	41637.87	0.22	0.641
Learning method:semantic congruency	1, 39	520.42	0	0.958
Learning method:gender congruency	1, 39	1023.38	2.37	0.132
Semantic congruency:gender congruency	1, 39	1147.49	0.08	0.78
Learning method:day	1, 39	1339.79	0.39	0.538
Semantic congruency:day	1, 39	921.13	0.22	0.643
Gender congruency:day	1, 39	1273.88	0.03	0.861
Learning method:semantic congruency:gender congruency	1, 39	1520.79	0.97	0.331
Learning method:semantic congruency:day	1, 39	899.15	0.01	0.942
Learning method:gender congruency:day	1, 39	1064.78	2.93	0.095
Semantic congruency:gender congruency:day	1, 39	1082.56	0.98	0.328
Learning method:semantic congruency:gender congruency:day	1, 39	1226.9	0.01	0.907

Study 3				
Learning method	1, 34	2436.95	0.64	0.429
Semantic congruency	1, 34	2057.7	0	0.974
Gender congruency	1, 34	2431.38	8.03	* .008
Day	1, 34	48176.61	0.9	0.35
Learning method:semantic congruency	1, 34	1338.71	8.58	* .006
Learning method:gender congruency	1, 34	1812.75	1.51	0.228
Semantic congruency:gender congruency	1, 34	1915.8	0.34	0.562
Learning method:day	1, 34	2452.28	0.71	0.405
Semantic congruency:day	1, 34	1051.73	0.37	0.544
Gender congruency:day	1, 34	2011.5	0.74	0.396
Learning method:semantic congruency:gender congruency	1, 34	3711.65	1.31	0.26
Learning method:semantic congruency:day	1, 34	1716.74	3.2	0.083
Learning method:gender congruency:day	1, 34	1903.39	0.88	0.356
Semantic congruency:gender congruency:day	1, 34	1163.92	0.55	0.464
Learning method:semantic congruency:gender congruency:day	1, 34	2028.4	0.03	0.861
Study 4				
Learning method	1, 31	1581.11	0.04	0.85
Semantic congruency	1, 31	803.84	2.75	0.108
Gender congruency	1, 31	1534.28	23.78	* <.001
Day	1, 31	30452.12	0.34	0.566
Learning method:semantic congruency	1, 31	541.28	0.02	0.885
Learning method:gender congruency	1, 31	809.34	1.94	0.174
Semantic congruency:gender congruency	1, 31	1048.71	1.86	0.183
Learning method:day	1, 31	887.78	0.02	0.892
Semantic congruency:day	1, 31	880.25	0.62	0.437
Gender congruency:day	1, 31	673.94	6.15	* .019
Learning method:semantic congruency:gender congruency	1, 31	856.99	0.05	0.83
Learning method:semantic congruency:day	1, 31	939	0.64	0.429
Learning method:gender congruency:day	1, 31	968.18	1.1	0.302
Semantic congruency:gender congruency:day	1, 31	1785.29	0.12	0.733
Learning method:semantic congruency:gender congruency:day	1, 31	559.84	0.12	0.734
Study 5				
Semantic congruency	1, 40	1857.58	0.02	0.893
Gender congruency	1, 40	1943.75	6.94	* .012
Day	1, 40	34602.71	2.01	0.164
Semantic congruency:gender congruency	1, 40	877.65	0.68	0.415
Semantic congruency:day	1, 40	2208.66	3.8	0.058
Gender congruency:day	1, 40	1697.69	2.72	0.107
Semantic congruency: gender congruency:day	1, 40	919.72	0.46	0.504

APPENDIX G

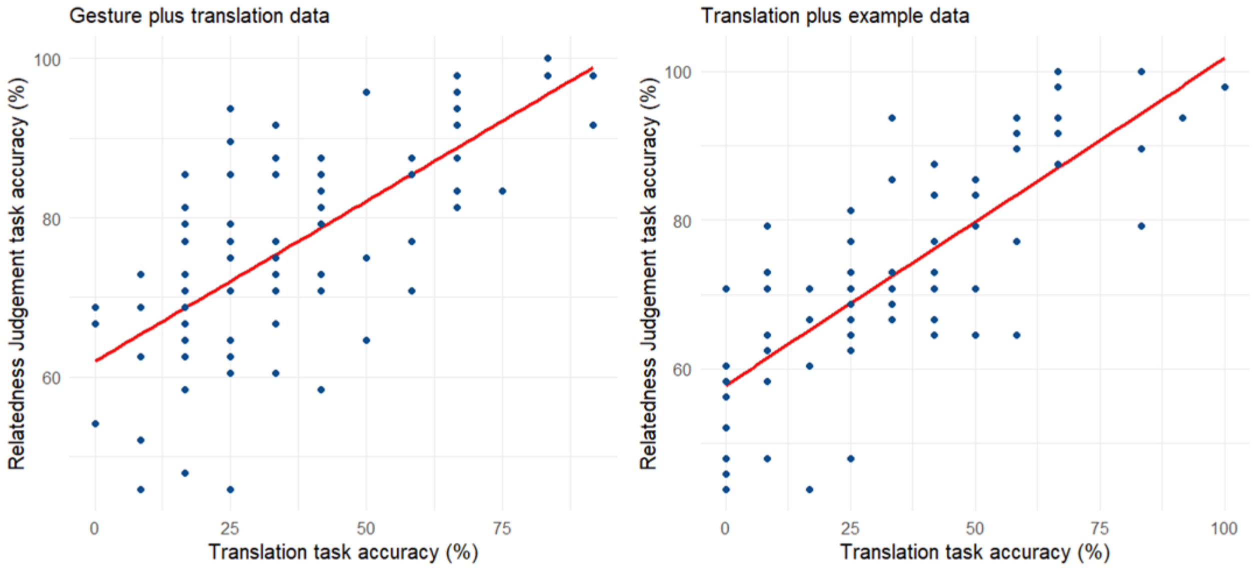
Analyses of the relationship between the 2 outcomes measures of translation accuracy & relatedness accuracy

Graph displaying the correlation between the relatedness judgment task percentage accuracy and the translation task percentage accuracy.



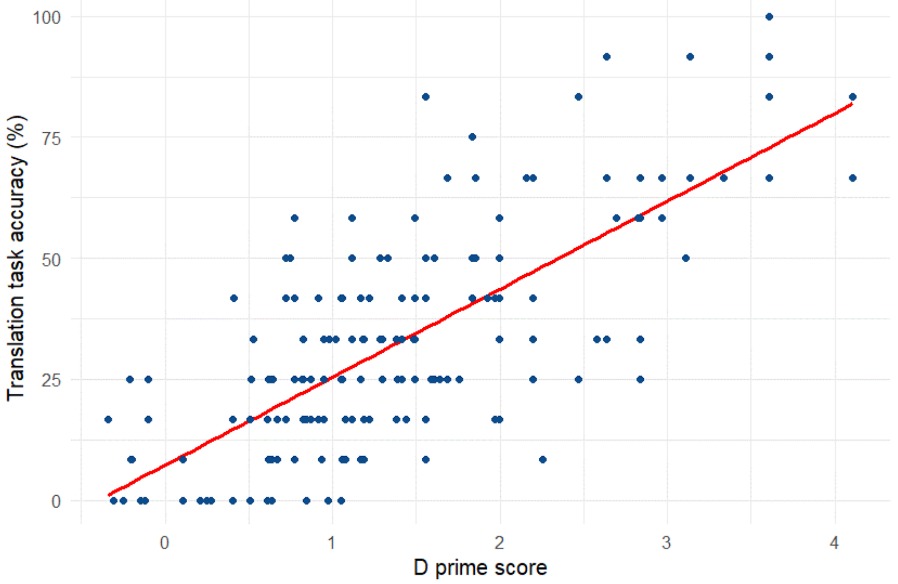
Note. There was a strong positive correlation between the percentage accuracy in the relatedness judgement task and the percentage of correct answers in the translation task, $r(174) = .74$, $p < .001$

Graphs displaying the correlation between the relatedness judgment task percentage accuracy and the translation task percentage accuracy for both learning methods separately.



Note. For the gesture plus translation learning method, a strong positive correlation was found, r(86) = .70, $p < .001$. A strong positive correlation was also found for the translation plus example learning method, r(86) = .78, $p < .001$.

Graph displaying the correlation between the translation task percentage accuracy and d prime scores

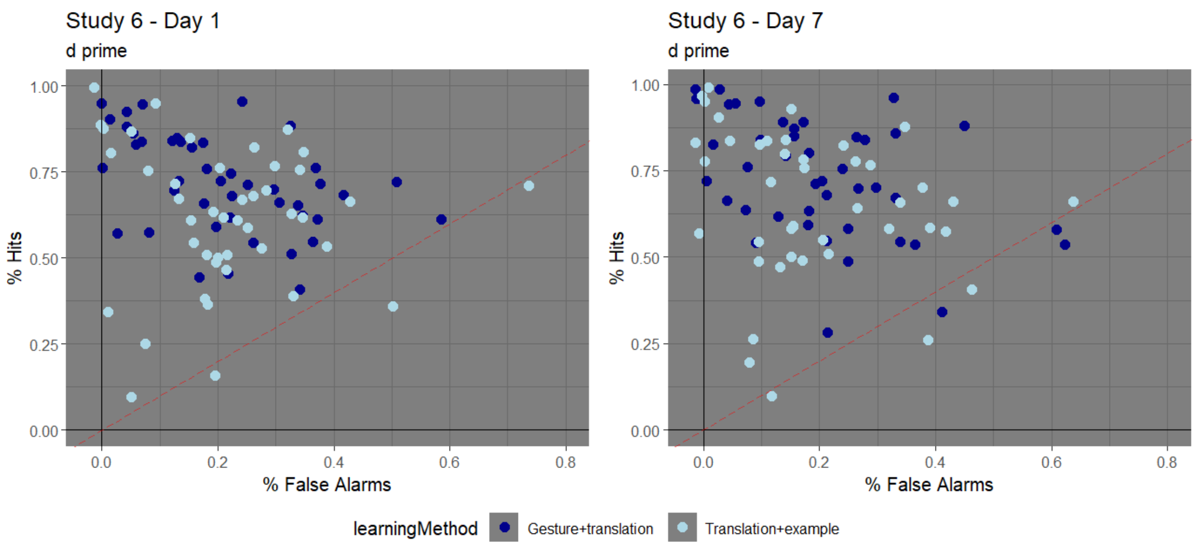


Note. A strong positive correlation was found r(174) = .75, $p < .001$

APPENDIX H

D-prime analysis for Study 6 without reaction time outliers removed

Graphs showing the proportion of false alarms and hits for day 1 & day 7 across the two learning methods



Note. The red dashed line indicates a d-prime of 0 (responses most closely associated with chance). A 2 (learning method) x 2 (day) ANOVA on unfiltered d-prime data was conducted. Significant effect of learning method, $F(1,43) = 7.57$, $p = .009$, significant effect of day, $F(1,43) = 5.91$, $p = .019$, no significant interaction between learning method and day, $F(1,43) = 0.03$, $p = .874$.

Graph showing the proportion of false alarms and hits for day 1 & day 7 across the two learning methods

Note. Graph displaying the d-prime scores across assessment day 1 and 7 for words learnt through gesture plus translation cues and translation plus example cues. The within-subject error bars reflect the variability of the data using 95% confidence intervals.

APPENDIX I

Participants' enactment of gestures during learning from end of study surveys

Study 2 responses

Participant	Enacted gesture?
1	yes
2	no
3	yes
4	no
5	no
6	yes
7	yes
8	no
9	yes
10	yes
11	yes
12	yes
13	yes
14	yes
15	yes
16	no
17	no
18	sometimes
19	yes
20	yes
21	sometimes
22	yes
23	yes
24	yes
25	yes
26	no
27	yes
28	sometimes
29	no
30	yes
31	sometimes
32	sometimes
33	yes

34	yes
35	yes
36	yes
37	yes
38	no
39	yes
40	yes
41	yes
42	no
43	yes
44	no
45	yes
46	yes
47	yes
48	yes
49	no
50	yes
51	no
52	no
53	no
54	yes
55	yes

Note. 63.64% of the respondents enacted the gestures, 33.85% did not and 8.8% sometimes enacted the gestures.

Study 4 responses

Participant	Enacted gesture?
1	yes
2	yes
3	no
4	yes
5	sometimes
6	sometimes
7	sometimes
8	no
9	yes
10	yes
11	sometimes
12	no

13	no
14	sometimes
15	sometimes
16	no
17	sometimes
18	yes
19	no
20	no
21	no
22	yes
23	sometimes
24	yes
25	yes
26	yes
27	yes
28	no
29	no
30	no
31	no
32	internally
33	yes
34	no
35	no
36	no
37	no
38	no
39	no
40	no
41	sometimes
42	yes
43	no
44	yes
45	yes

Note. 33.33% of the respondents enacted the gestures, 46.67% did not and 20% sometimes enacted the gestures.

Study 5 responses

Participant	Enacted gestures?
1	yes
2	sometimes
3	yes
4	no
5	no
6	yes
7	no
8	no
9	no
10	no
11	no
12	yes
13	no
14	sometimes
15	no
16	yes
17	yes
18	yes
19	no
20	yes
21	yes
22	no
23	yes
24	no
25	yes
26	yes
27	yes
28	no
29	yes
30	yes
31	yes
32	yes
33	yes
34	no
35	no
36	sometimes
37	no
38	yes
39	yes

40	no
41	no
42	no
43	sometimes
44	yes
45	yes
46	yes
47	yes

Note. 68.09% of the respondents enacted the gestures, 22.22% did not and 10.64% sometimes enacted the gestures.

Study 6 responses

Participant	Enacted gestures?
1	no
2	no
3	no
4	Yes
5	no
6	yes
7	yes
8	sometimes
9	sometimes
10	yes
11	sometimes
12	no
13	no
14	no
15	yes
16	no
17	yes
18	yes
19	no
20	no
21	yes
22	no
23	yes
24	sometimes
25	no
26	no

27	no
28	no
29	no
30	sometimes
31	yes
32	yes
33	no
34	no
35	yes
36	yes
37	no
38	yes
39	yes
40	no
41	yes
42	no
43	yes
44	no

Note. 38.64% of the respondents enacted the gestures, 50% did not and 11.36% sometimes enacted the gestures.

APPENDIX J

Self-reported English proficiency scores for non-native pre-test study

Participant	Gender	Age	Reading	Understanding	Speaking	Language spoken at home
1	M	20	10	10	10	other
2	F	30	10	10	10	both equally
3	F	27	9	8	9	other
4	M	28	9	9	7	other
5	M	32	10	9	7	other
6	F	22	10	8	8	other
7	F	24	10	10	10	other
8	M	41	9	8	7	other
9	F	26	8	7	7	other
10	F	38	9	10	10	other
11	M	22	10	8	7	other
12	M	20	9	10	9	other
13	M	20	9	7	5	other
14	F	41	8	8	9	other
15	M	20	9	8	7	other
16	M	22	8	9	7	other
17	F	27	9	8	8	other
18	F	31	10	9	9	English
19	M	63	7	6	6	other
20	M	36	10	8	9	other
21	F	24	10	10	10	other
22	F	28	9	9	9	other
23	F	65	7	6	6	other
24	F	25	9	8	8	other
25	F	50	8	6	6	English
26	F	25	8	9	8	other
27	F	20	10	10	10	other
28	F	20	10	9	8	English
29	F	25	8	7	7	other
30	M	18	9	10	10	other
31	M	27	7	7	7	other
32	M	20	10	10	9	other
33	F	20	10	10	10	other
34	F	20	10	10	10	other
35	M	19	5	5	5	other
36	F	20	10	10	10	other
37	F	20	10	10	10	both equally
38	F	21	9	7	9	other
39	M	22	9	9	9	English
40	M	21	8	8	8	both equally