This is a pre-publication version of Rodríguez-Cuadrado, S., Ojedo, F., Vicente-Conesa, F., Romero-Rivas, C., Sampedro, M. A., & Santiago, J. (accepted). Sign iconicity helps learning new words for abstract concepts in a foreign language. *Second Language Research.*

Sign iconicity helps learning new words for abstract concepts in a foreign language

Sara Rodríguez-Cuadrado[1*]; Fernando Ojedo[2]; Francisco Vicente-Conesa[3]; Carlos Romero-Rivas[1]; Miguel Ángel Sampedro[4]; & Julio Santiago[2]

[1]Departamento Interfacultativo de Psicología Evolutiva y de la Educación, Universidad Autónoma de Madrid,

[2]Departamento de Psicología Experimental, Universidad de Granada,

[3]Departmento de Psicología Básica, Universidad Autónoma de Madrid,

[4] Facultad de Lenguas y Educación, Universidad Nebrija

*Corresponding autor:
Sara Rodríguez-Cuadrado
Facultad de Formación del Profesorado y Educación,
Universidad Autónoma de Madrid
C/ Francisco Tomás y Valiente, n° 3, 28049, Madrid
sara.rodriguezcuadrado@uam.es

Abstract

Several studies have explored the use of iconic gestures to improve the learning of foreign vocabulary. In this quest, words for abstract concepts have been largely neglected, under the assumption that abstract concepts have poor or inexistent sensory-motor representations. Yet, the Conceptual Metaphor Theory suggests that they are grounded on concrete concepts. Moreover, analyses of signed languages reveal ways in which signs can exploit metonymies and conceptual metaphors to iconically refer to abstract concepts. Here, we explore whether iconic signs from Spanish Sign Language (LSE) can facilitate the learning of foreign words for abstract concepts in hearing participants who do not know any sign language. In two studies, participants were presented with new labels for abstract and concrete concepts in an artificial language (Vimmi). The labels could be accompanied by either a video of an iconic or non-iconic sign taken from the existing vocabulary of LSE, or a static image of the signer. In study 1, participants did not have to enact the signs they were presented with, while in study 2 they did. Both studies showed that iconic signs facilitated the learning of abstract foreign vocabulary, regardless of enactment. The strategies that sign languages use to develop iconic signs for abstract concepts make those signs useful to assist the learning of foreign words by hearing non-signers.

Keywords: abstract words, iconicity, sign language, Conceptual Metaphor Theory, vocabulary learning, enactment

1. Introduction

Over the past few years, a prolific body of research has investigated the role of iconic gestures in foreign vocabulary learning (FVL). Iconicity refers to a relation of resemblance between the form of the communicative signal and the sensory-motor characteristics of its referent (Dingemanse et al., 2015). In the case of expressive gestures of the hands and face, iconicity is motivated by visual-spatial correspondences between gesture and referent (Ortega, 2017). In the case of vocalizations, iconicity is motivated by correspondences between aspects of the auditory signal and aspects of the referent (e.g., high pitch is iconic of a small-sized object; Knoeferle et al., 2017). Conventional signals, that is, lexical items from natural languages, can also be iconic (see, e.g., Miozzo et al., 2020, for words; Blasi et al., 2016, for signs). Iconic lexical items (either signs or words) are constrained by the structure of the language and follow language-dependent phonological principles (see Ortega, 2017), while expressive gestures are not.

Iconicity is important in language acquisition. It positively influences language learning during infancy (Imai and Kita, 2014; Massaro and Pearlman, 2017; Perniss and Vigliocco, 2014; Perry et al., 2018; Sidhu et al., 2022), and bears a clear relationship with age of acquisition in spoken (e.g., Hinojosa et al. 2020; Perry et al., 2015) and signed languages (e.g., Caselli and Pyers, 2017; Thompson et al., 2012; Vinson et al., 2008). The learning benefit offered by iconic words, gestures and signs has been mostly explained by the embodiment hypothesis, for which cognition is rooted in perception and action (Meteyard et al., 2012), emphasising how high-level mental processes (such as language processing) are related to the body (Barsalou et al., 2003).

Iconicity relies on concreteness. Concreteness is a property of concepts. Concrete concepts have referents that can be perceived (i.e., seen, touched) and/or interacted with. These referents are endowed with "physical appearance" (broadly understood as a set of associated specific perceptual and motor experiences). Their physical appearance can be used to guide the formation of iconic signals, which are then used to communicate about those referents. Applied to vocabulary learning, iconic lexical items cue learners to recreate the sensory-motor features of the referents and, therefore, they should be learnt more easily when first encountered than purely arbitrary lexical items (see Macedonia, 2014). If iconicity helps first language vocabulary learning, it is to be expected that it may also help FVL. Moreover, when the new lexical item to be learnt is not iconic, its degree of iconicity can be boosted by presenting it together with an iconic picture or gesture. Intuitively, iconic gestures should make an even stronger connection with the referent than pictures, as they implicate the body to a greater extent. Below we will review the available evidence on the question of whether pairing new arbitrary verbal labels (vocabulary items of foreign languages) with pictures and gestures that are iconic of the concrete referents of the label facilitates the long-term learning of the association between the new label and its meaning.

Yet, the expectation of positive effects of iconicity on FVL breaks down when it comes to the learning of lexical items that refer to abstract concepts such as “belief” or “time”. By definition, abstract concepts have referents that cannot be seen or touched. Therefore, it is difficult to see how they can be grounded on sensory-motor experiences. In fact, abstract concepts are recognized as posing a formidable challenge to embodied cognition accounts (Barsalou, 1999; Borghi et al., 2017; Dove, 2015). Applied to FVL, the learning of new labels for abstract concepts cannot, in principle, take advantage of an iconic connection between the label and its meaning. As a communicative signal cannot resemble an abstract referent, the form of the signal cannot help the learner to recreate any physical appearance of the referent and thereby assist them in learning the association between signal and referent. Consequently, only a handful of studies have used gestures to teach new words for abstract concepts (Andrä et al., 2020; Macedonia and Knösche, 2011; Mayer et al., 2017; Repetto et al., 2017). Those studies have all relied on experimenter intuition to generate gestures that might possibly be helpful for learning, with some positive results (see below). However, they cannot offer a principled account of why those gestures help the learner establish the connection between the signal and its meaning.

In the current article we take a principled approach to understanding how a communicative signal can be made iconic of an abstract concept. We work from a theoretical approach that defends that abstract concepts get grounded into concrete concepts through “imaginative” processes of the human mind, including conceptual metaphor and metonymy (Lakoff and Johnson, 1980; Lakoff et al., 1999; to be spelled out below). Through these mental operations, abstract concepts are understood with the help of concrete concepts, which in turn have referents that can be seen and touched. Therefore, signals for abstract concepts can be made iconic to the physical appearance of those referents. When applied to the analysis of the vocabulary of sign languages, this approach affords the distinction between iconic and arbitrary signs for abstract concepts (Taub, 2001; Wilcox, 2000). Here we will test whether the iconicity of new words for abstract concepts can be boosted by pairing them with iconic signs from the existing vocabulary of sign languages and, thereby, help hearing people who do not know any sign language to learn those words. Our main departure from standard practices in this field of research is, therefore, that instead of using non-conventionalised communicative gestural signals, we use highly conventional hand actions, i.e., lexical signs of a sign language.

In the following lines, we will first review past work on FVL which has examined the effects of enriching the memory trace of the new vocabulary items by means of their pairing with pictures and gestures on the learning of new vocabulary for concrete and/or abstract concepts. We will then spell out in detail how signs of sign languages can be iconic of concrete referents and, more importantly, how they can be iconic of abstract referents.

1.1. Foreign vocabulary learning as paired associate learning

In studies of FVL, participants are typically presented with pairs of an L1 word and a new word in an unknown language (L2), often an artificial language such as Vimmi to control for prior knowledge and similarities to L1 (see Ettlinger et al., 2016, for a validation of this strategy). The task is to learn the new labels for the already known words and meanings. This learning is often assessed by means of cued recall from L2 into L1, although other measures such as cued recall from L1 into L2 and free recall are also common. Participants usually go through several learning and testing sessions.

Important research questions in this field have been whether labels of concrete concepts are easier to learn than labels of abstract concepts and whether learning can be boosted by enriching the memory trace by adding an additional stimulus to the words. As mentioned above, the hypothesis of embodiment has motivated special attention to the potential beneficial effects of a particular kind of stimuli: gestures. Gestures are communicative actions that involve the hands and the face. Therefore, gestures possibly implicate the body in the learning task to a greater extent than, say, a picture or another word. Thus, the question of enrichment has branched into several questions: whether gestures are more effective than pictures as additional stimuli; whether enacting gestures provides an extra learning benefit over just observing them; and finally, whether gestures exert their effect due to their iconic connection with the word meaning. Available studies have used designs that tap into these questions in partial ways, for example, comparing a no enrichment condition with a condition where words for abstract concepts are enriched with iconic enacted gestures (Macedonia and Knösche, 2011). Thus, clear answers regarding the main effects and interactions of concreteness, gesture enrichment, iconicity, and enactment are unavailable, as all studies leave unexplored some combinations of these factors.

In general, the concreteness effect seems well established: learning new labels for concrete concepts is easier than for abstract concepts (Andrä et al., 2020; Kaushanskaya and Rechtzigel, 2012; Mayer et al., 2017). This is in line with the well-known effect of concreteness on lexical processing and memory (e.g., Paivio et al., 1994). In contrast, disagreements abound regarding the effect of enrichment by means of pictures. Enriching the memory trace leads to better learning and recall, as predicted by the Dual Coding theory (Clark and Paivio, 1991), but mainly when it implies an active processing of the material (e.g., by imagining a complex scene in which the learner interacts with the material, as in Van Hell and Mann, 1997). Studies that enrich the memory trace by passively viewing pictures have reached intricate results. Mayer et al. (2015) and Morett (2019) observed better learning with pictures than without enrichment, but Rowe et al. (2013) found that pictures affected only children that were learning two languages (versus one), helping those with low language abilities and hindering those with high language abilities. Repetto et al. (2017) also observed learning benefits of picture enrichment only in some measures, and often the by-subjects and by-items analyses did not converge.

It is also quite unclear whether FVL for concrete concepts benefits more from enriching with iconic gestures than with pictures, as there are frequent problems and contradictory findings in the literature. Tellier (2008) found better learning with enacted gestures than with pictures in French children learning English, and Porter (2016) found the same result in English children learning French. Unfortunately, both studies suffer from too small samples of participants and items, among other problems. Andrä et al. (2020, exp. 3) found similar Vimmi learning with pictures and enacted gestures, but they lacked a no enrichment condition, so it is impossible to discern whether both conditions had the same effect or had no effect at all. Mayer et al. (2015) reported that enacted gestures improved Vimmi learning more than enacted pictures (following their outline with the finger), but when both gestures and pictures were passively viewed, pictures outperformed gestures. Morett (2019) also reported the same pattern: picture observation improved Hungarian word learning more than gesture observation in English adults.

Evidence is also ambiguous regarding whether enrichment by means of iconic gestures (even when enacted) is superior to no enrichment at all in FVL of words for concrete concepts. In two studies, Morett (2018, 2019) failed to observe any learning benefit of iconic gesture observation, while Macedonia et al. (2019) only found benefits in free recall at the short-term, but not in paired and cued recall, either at the short or long term. Macedonia et al. (2014) failed to find significant learning benefits for gesture observation and gesture enactment over no enrichment in German children learning English with the help of a virtual avatar (this description of their findings follows from the figures because there are issues with the statistics reported in the paper). Using enacted iconic gestures, Macedonia et al. (2010, 2011), Macedonia and Klimesch (2014) and Mayer et al. (2015) reported a facilitatory effect of gesture enrichment over no enrichment, but neither Andrä et al. (2020) nor Krönke et al. (2013) found this effect. One additional study suggests that there may be mediating factors: Kroenke et al. (2013) failed to find an overall learning benefit of enacting gestures in their mild-aphasic sample but a sub-class of patients with lesser lexico-semantic impairments did benefit from gesture enrichment.

Findings seem more consistent with an effect of gesture enrichment when we look at studies that include incongruent gestures (gestures which are iconic of other meanings in the learning set). In this case, both Kelly et al. (2009) and Garcia-Gamez and Macizo (2019) reported that congruent gestures facilitate learning whereas incongruent gestures interfere with learning with respect to a no gesture condition. This is interesting because it supports that the effect of gesture enrichment is caused by the iconic link between the gesture and the meaning of the new label, and not just by an increase in attention, arousal, mere stimulus complexity or the addition of motor behaviour. However, when iconic gestures are compared with meaningless gestures, some studies find that the latter hinder learning (Garcia-Gamez and Macizo, 2019; Macedonia et al., 2010), others find that they produce the same learning as control conditions (Macedonia et al., 2011), while still others find that they facilitate learning to the same extent as iconic gestures (Huang et al., 2019). It is also important to point out that, in these studies, the inclusion of incongruent and meaningless gestures may favour learning strategies that rely more strongly on gesture iconicity in the congruent condition. Words learned with gestures (enacted or not) produce different patterns of brain activity than words learned without gestures (Kelly et al., 2009; Macedonia et al., 2010, 2011, 2019; Macedonia and Mueller, 2016; Mayer et al., 2015; but see Mayer et al., 2017, for null results), but so far it is unclear whether this translates into behavioural differences.

In the same vein, the evidence supporting an effect of enrichment when learning labels for abstract concepts is also far from convincing. Mayer et al. (2017) compared learning Vimmi words for abstract concepts with no enrichment, picture or gesture enrichment, and reported their results collapsing one experiment in which participants enacted the gesture and the picture (following its outline with the finger) with another experiment without enactment. They found no significant effect of any kind of enrichment neither for concrete nor abstract concepts. In contrast, Andrä et al. (2020) presented English words referring to both concrete and abstract concepts to German eight-year-old children without enrichment or with enacted gestures in two experiments, and observed only in the first experiment a learning benefit with gesture enrichment and only for the abstract concepts. Both Macedonia and Knösche (2011) and Repetto et al. (2017) found that the learning of Vimmi labels for abstract concepts benefited from gesture enactment over no enrichment. However, neither study is informative regarding the interaction of gesture enrichment with concreteness:

Repetto et al. (2017) because they only used abstract concepts, and Macedonia and Knösche (2011) placed all concrete nouns as subjects and all abstract nouns as objects in sentences, making it difficult to interpret any differences between them. Repetto et al. (2017) also provide unclear findings, often showing only in the by-subjects analysis and not in the by-item analysis, and in some measures but not in others.

Finally, whether enacting the gesture produces better learning than passive observation of the gesture is also still an open question. Direct comparisons between those two conditions have been carried out in only three studies. Macedonia et al. (2014) found similar levels of FVL in the two enrichment conditions and in the no gesture condition (judging from their figures); Krönke et al. (2013) also failed to find any difference between enactment and passive viewing; and de Nooijer et al. (2014) found learning benefits of gesture enactment over mere gesture observation only for object-manipulation verbs but not for locomotion or abstract verbs. Morett (2018) presented gestures for observation while her participants learnt foreign words and let them gesture spontaneously while they explained the new words to another participant. Thus, the observed and produced gestures were not necessarily the same, complicating the interpretation of her data.

All in all, only the concreteness effect is clearly established in the field of FVL. Despite the accumulation of a sizeable literature, no clear conclusions can be reached regarding the effects on the learning of new words for concrete versus abstract concepts, with enrichment versus without enrichment, by pictures versus gestures, under passive viewing versus enactment conditions, iconically related versus unrelated to the lexical meanings. The present study aims to shed some light on these questions by assessing into a single design the learning of new labels for concrete and abstract concepts presented together with manual movements (signs from LSE, *Lengua de Signos Española* or Spanish Sign Language) that are either iconically or arbitrarily related to the word meanings, as well as a visual control condition (a still picture of the signer), and testing this design in two experiments differing only on whether the signs are enacted or not. But before, we need to establish how signs from sign languages can be iconic of concrete and, specially, abstract concepts.

1.2. Iconicity in sign language vocabulary

Taub (2001) proposed an "analogue-building model" of linguistic iconicity which suggests that iconic lexical items, both oral and signed, are created through a three step process: first, a mental image of the concept is selected; second, the image is schematised (i.e., central aspects of the image are selected); and finally, phonological components of the language are selected and combined to assemble a form that retains the same structure of the schematic image while complying with linguistic constraints. Iconicity is pervasive in sign languages, while relatively sparse in oral languages, because linguistically licensed hand shapes and movements afford much greater possibilities than sounds to enter into configurations that respect the structure of the schematic image (Taub, 2001; Wilcox, 2000).

The LSE sign for "gorilla" may serve as an example (see Figure 1). Out of the many images related to a gorilla, LSE selects the image of a male gorilla beating his chest. The image is schematised to include only the alternating movements of the fists pounding on the chest. Finally, a sign that respects this schematic structure is composed using a linguistically sanctioned handshape and movement. As shown by the gorilla sign, iconic signs for concrete concepts often have a strong metonymic component (Sutton-Spence and Woll, 1999): in this case, a characteristic movement of male gorillas is used to stand for the whole category.

Figure 1. *The "gorilla" sign in Spanish Sign Language (LSE).*

[INSERT FIGURE 1]

Besides metonymy, signs for abstract concepts also take advantage of conceptual metaphors. These are operations of the mind that allow us to understand a concept in terms of another (Lakoff and Johnson, 1980; Lakoff et al., 1999). For example, the concept of "time" is understood as movement in space from a location (in the past) to another location (in the future) along a path, as shown by sentences like "we left those times behind". Following Taub (2001), once a conceptual metaphor (or several, see below) establishes a correspondence between an abstract and a concrete concept, an iconic sign for that concrete concept can be created following the same three step procedure, with the conceptual metaphor guiding the selection and schematisation of the image. The sign for "belief" (Figure 2) illustrates how metonymies and conceptual metaphors are often invoked together to create an iconic sign for an abstract concept. The handshape of the "belief" sign closes the thumb and index fingers into a circle. The tip of those fingers are brought to the temple, as the head metonymically stands for mental activity (many other signs of mental activities, such as "idea" or "remember", are also articulated at the head). The hand then travels down to central signing space and stops like posing the hand onto a solid surface. The sign iconically invokes two conceptual metaphors, one that conceives of thoughts as buildings and another that likens certainty to solidity (manifested in oral languages in sentences such as "that theory has solid foundations").

Figure 2. *The "belief" sign in Spanish Sign Language (LSE).*

[INSERT FIGURE 2]

Taub (2001) and Wilcox (2000), among others, have shown that it is possible to successfully analyse many signs of abstract concepts in terms of metonymies and conceptual metaphors. Those signs would, therefore, be considered iconic of the referents of those abstract concepts (actually, they are iconic of the concrete concepts which help grounding those abstract concepts). Iconic signs can be contrasted with arbitrary signs, signs which either never had or have lost their iconicity due to progressive simplification and stylisation with the historical change of the language (see, e.g., Frishberg, 1975). Despite rampant iconicity, it is also easy to find arbitrary signs for both concrete and abstract concepts in sign languages. In the present study we lean on this property of the vocabulary of sign languages to build our materials.

1.3. The present research

Our main research question is whether iconic signs from sign languages can help provide a principled understanding and a systematic procedure to teach foreign vocabulary for abstract concepts. Specifically, we tested whether signs from LSE that are iconic of the concrete grounding of abstract concepts can help the learning of new arbitrary verbal labels for those concepts over and above the increase in richness of the memory trace due to the presentation of the label together with an arbitrary gesture (either seen or performed). Additionally, we sought to replicate three important effects in this field (see review above) using signs instead of gestures: 1) the effect of gesture iconicity for concrete concepts; 2) the effect of gestural enrichment, that is, the learning benefit due to presenting a new label together with a gesture versus a still picture of the same person; and 3) the effect of enactment, that is, the learning benefit produced by watching plus enacting the gesture over only watching it. To do so, we compared the learning of new labels in a foreign (artificial) language for both concrete and abstract concepts conveyed through their L1 labels when these were paired with either iconic or arbitrary LSE signs as well as in a no sign condition (a still picture of the signer). This design was implemented in two experiments, one with instructions of only watch the sign (Exp. 1) and another with instructions to watch and enact the sign (Exp. 2).

Based on the theoretical background offered by the Conceptual Metaphor Theory (Lakoff and Johnson, 1980; Lakoff et al., 1999) and its application to the analysis of sign languages (Taub, 2001; Wilcox, 2000), we hypothesise that signs from existing sign languages should be an effective tool to teach foreign abstract vocabulary to hearers who do not know those sign languages. Signs could provide the most benefit for abstract words as they are more difficult to learn to begin with. Additionally, we also explore whether iconic signs can facilitate FVL of concrete concepts; and, by comparing studies 1 and 2, whether the enactment of the accompanying sign provides an additional benefit over its passive viewing. As far as we can tell, only one study has posed similar questions to the present one, but it has done so regarding the learning of the signs themselves. Morett (2015) compared how well non-signers learn arbitrary signs (most of them with abstract referents) and iconic signs of concrete and abstract concepts (that she calls "metaphorics"). She found that the three kinds of signs were equally learnt after a single session of training, but that in the long-term learning was better for the concrete iconic, followed by the abstract iconic, and finally by the arbitrary signs. As the present study applies these sign distinctions to enriching the learning of oral labels, we are here researching unexplored terrain.

2. Study 1

2.1. Method

2.1.1. Participants

Thirty-nine students from the University of Granada and the University of Murcia (mean age = 20.55, *SD* = 2.34, five males, all right-handed) took part in the study, and were compensated with course credits. All participants were Spanish native speakers and had no prior knowledge of any sing language. All had normal or corrected-to-normal vision and no hearing impairments. All participants provided signed consent prior to participating in the research, and were given an information sheet on the first session and a debriefing sheet on the last session. Both sheets informed the participants of their right to withdraw from the study as well as about the protection of anonymity and confidentiality of the collected personal information. The study was conducted in accordance with the ethical standards laid down in the Declaration of Helsinki and was approved by the ethics committee of Edge Hill University's Department of Psychology.

2.1.2. Materials

Each participant was presented with 48 words in Spanish and their Vimmi translation. Half of these words referred to concrete and half to abstract concepts. The words were accompanied either by a static image of the signer ("no sign" condition, 16 words), or by a video of the signer performing an iconic (16 words) or a non-iconic (16 words) sign. Within each concrete or abstract condition, eight words were presented in the "no sign", eight in the "iconic", and eight in the "non-iconic" conditions. There were six different lists of Spanish-Vimmi word pairs, so each Vimmi word was paired to a Spanish word belonging to each one of the six possible word-sign pairs: abstract/iconic (AI), abstract/non-iconic (ANI), abstract/no sign (ANS), concrete/iconic (CI), concrete/non-iconic (CNI), and concrete/no sign (CNS; to see randomisation lists, see Tables 3-8 in the Appendix available at the OSF project; see below). Length of Spanish and Vimmi words was matched so length did not vary between languages ($t(94) = -.215$, $p = .83$) or conditions ($F(5) = .039$, $p = .99$).

The signer in our videos is a deaf LSE linguist and teacher (and a team member). He categorised the signs as iconic or non-iconic, and their concepts as concrete or abstract. Iconicity was also assessed independently by hearing non-signers, who rated a subset of randomly selected iconic and non-iconic signs in a 7-point Likert scale (1 = completely non-iconic to 7 = completely iconic). Mean difference between iconic and non-iconic signs was 2.91, and a related samples t-test showed that the difference was significant ($t(10) = 4.74$, $p = .001$). Therefore, the ratings by hearers supported the classification of signs as iconic or non-iconic made by our LSE specialist. Concreteness was also assessed independently, by retrieving their values from databases by Guasch et al. (2016) and Duchon et al. (2013). Mean difference between words for concrete and abstract concepts was 2.08 (in a 7-point Likert scale), and an independent samples t-test showed a clear difference ($t(29{,}801) = 8.36$, $p < .001$). Thus, the classification of concepts as concrete or abstract made by our LSE specialist was also independently supported. All signs were monomorphemic.

Spanish (L1) words were presented visually (see Table 1 in the Appendix available at the OSF project; see below). Only nouns were included to eliminate potential confounds and to maximise word learning (Macedonia and Knösche, 2011; Marshall, 2003; Waxman et al., 2013). Lexical values including number of letters, logarithmic frequency, and word class were retrieved from the NIM database (Guasch et al., 2013).

Forty-eight Vimmi (artificial L2; Macedonia et al., 2011; Macedonia and Knösche, 2011) words were presented both visually and aurally (see Table 2 in the Appendix available at the OSF project; see below). We chose Vimmi because it allowed us to control for any prior foreign language knowledge, and because it has been used in previous language learning studies (Bergmann and Macedonia, 2013; Garcia-Gamez and Macizo, 2019; Macedonia et al., 2011, 2014; Macedonia and Knösche, 2011; Macedonia and Mueller, 2016).

Vimmi words were recorded by a male native Spanish speaker, cut into individual audiofiles, and balanced in terms of intensity (dB), normalising the maximum amplitude across all experimental words using the software Audacity®. Each audiofile lasted approximately one second.

In the videos, the signs corresponding to the Spanish words were produced by the LSE male signer. Each participant was presented with 32 videos, eight per condition (AI, ANI, CI, CNI). For the no sign condition, participants were presented with a static image of the signer (Figure 3) 16 times (eight times for the ANS condition, eight times for the CNS condition). Both the videos and the static image lasted approximately three seconds.

Figure 3. *Static image of the signer.*
[INSERT FIGURE 3]

2.1.3. Procedure

Participants were presented with three training blocks and three testing sessions on the same day (from now on, T1, T2 and T3), and a fourth testing session a week later (T4), which allows assessing the durability of the learning effects by exploring short-term and long-term memory (Morett, 2019).

The experiment was programmed and presented using E-Prime 2.0 Professional (Schneider et al., 2007) and administered individually in a quiet experimental booth. To ensure that participants engaged in the task, one of the experimenters sat in the same booth during training (but not during testing), out of sight to prevent distractions.

Participants were informed that the experiment consisted of learning foreign nouns, and that they will be tested after each of the three training blocks (T1, T2, T3) and once again one week later (T4), so they should try to learn as many words as possible. On the first session, before starting with the experimental blocks, a small practice with three items was provided to ensure that participants understood the task. Each block contained 48 words and videos/static images, which were presented randomly. Conditions were mixed within each block, and not presented in separate blocks to avoid participants to notice our experimental manipulations.

Trials had a learning and a repetition phase and were as follows: First, a red cross was presented in the middle of the screen for 1000 ms, signaling the "learning" phase of the trial. Participants were instructed to only listen and watch at this stage. Then, the word in Vimmi was presented and the video (or static image) was shown for a total of 3500 ms. As the video faded out, the translation into the L1 appeared. After 3000 ms, a green cross showed in the middle of the screen, signaling the "repetition" phase of the trial. For the repetition phase, the word in Vimmi was presented, the video or image appeared, and the translation emerged as the video faded out, for 3500 ms (see Figure 4). Participants were instructed to repeat the word out loud in Vimmi as it appeared on the screen in the repetition phase. Each training block lasted approximately 15 minutes.

Once block 1 was finished, participants completed a cued recall task (see below for details). Then, the learning and cued recall blocks were repeated two more times. Trial order was randomised in all blocks. A week later, participants returned to the lab to fill a final cued recall sheet. No training was delivered on that session.

Cued recall consisted of a L2 (Vimmi) to L1 (Spanish) translation task. Participants were presented with a sheet of paper including the 48 words in Vimmi in one column, and a blank column for Spanish translations. The order in which the Vimmi words were presented varied in all testing blocks, to control for learning position effects. The whole first session (T1, T2, T3) took approximately one hour. T4 took about 15 minutes. Participants' responses were categorised as "omission" (no translation provided), "hit" (correct translation) and "error" (incorrect translation). The number of hits was the dependent variable.

Figure 4. *Procedure followed in studies 1 and* 2.
[INSERT FIGURE 4]

2.2. Results

Results are organised as follows. We first present analyses comparing hits in the sign trials (including videos of both iconic and non-iconic signs) versus the no sign trials. This allows us to check whether enrichment by signs in general enhance FVL in comparison to only presenting verbal information for both abstract and concrete concepts. Then, analyses are performed comparing the type of sign (iconic versus non-iconic), to evaluate whether iconicity facilitates FVL for both abstract and concrete words.

Data were analysed with a logit-mixed model implemented in R (lme4 library; Bates, 2010). We included the variables *test time, concreteness,* and either *sign* or *iconicity* (and all the possible interactions between them) as fixed factors. The random term included random intercepts for both participants and items, as well as random slopes for test time, concreteness, and sign/iconicity over participants and items. For the sake of brevity, we only describe significant slopes and their corresponding post-hoc analyses in this section.

Full disclosure of fixed and random effects, together with estimates, standard errors, z scores and p values for the slopes for all the analyses reported here can be found in Tables 1 and 2 at the OSF project (see below). Number of hits was always the dependent variable. Figure 5 shows the main results.

Figure 5. *Recall in studies 1 and 2.*

[INSERT FIGURE 5]

Note. Figure 5 shows recall accuracy (from 0 to 1) for new words in the abstract iconic, abstract non-iconic, concrete iconic and concrete non-iconic conditions in studies 1 and 2. Error bars show standard error values.

2.2.1. Sign versus no sign

Independent variables were *test time* (T1, T2, T3, T4), *concreteness* (concrete versus abstract) and *sign* (sign versus no sign).

There was a main effect of test time (*Estimate* = 0.38, *SE* = 0.03, z = 11.03, $p < .001$). Post-hoc analyses showed higher accuracy in T2 versus T1 (M_{T1}= 0.07, SD_{T1} = 0.26, M_{T2} = 0.31, SD_{T2} = 0.46; *Estimate* = 1.06, *SE* = 0.08, z = 13.22, $p < .001$); higher accuracy in T3 versus T1 (M_{T3} = 0.51, SD_{T3} = 0.5; *Estimate* = 1.58, *SE* = 0.09, z = 16.89, $p < .001$); higher accuracy in T4 versus T1 (M_{T4} = 0.32, SD_{T4} = 0.46; *Estimate* = 1.04, *SE* = 0.09, z = 11.77, $p < .001$); higher accuracy in T3 versus T2 (*Estimate* = 0.54, *SE* = 0.05, z = 11.68, $p < .001$); no significant differences between T2 and T4 (*Estimate* = -0.01, *SE* = 0.04, z = -0.14, p = .89); and higher accuracy for T3 versus T4 (*Estimate* = -0.54, *SE* = 0.04, z = -12.72, $p < .001$).

We also observed a main effect of concreteness (*Estimate* = 0.24, *SE* = 0.65, z = 3.75, $p < .001$), with higher accuracy in the recall of concrete words (M = 0.34, SD = 0.47) than abstract words (M = 0.27, SD = 0.44).

In addition, an interaction between test time and concreteness was found (*Estimate* = 0.07, *SE* = 0.02, z = 3.34, $p < .001$). Post-hoc analyses revealed no differences in recall between concrete and abstract words at T1 ($M_{concrete}$ = 0.08, $SD_{concrete}$ = 0.27, $M_{abstract}$ = 0.07, $SD_{abstract}$ = 0.26; *Estimate* = 0.16, *SE* = 0.15, *z* = 1.03, *p* = 0.31). However, concrete words were more accurately recalled than abstract words at T2 ($M_{concrete}$ = 0.34, $SD_{concrete}$ = 0.47, $M_{abstract}$ = 0.29, $SD_{abstract}$ = 0.46; *Estimate* = 0.17, *SE* = 0.07, *z* = 2.47, *p* = .01), T3 ($M_{concrete}$ = 0.56, $SD_{concrete}$ = 0.50, $M_{abstract}$ = 0.46, $SD_{abstract}$ = 0.50; *Estimate* = 0.31, *SE* = 0.08, *z* = 4.02, *p* < .001) and T4 ($M_{concrete}$ = 0.37, $SD_{concrete}$ = 0.48, $M_{abstract}$ = 0.27, *SD* = 0.44; *Estimate* = 0.35, *SE* = 0.07, *z* = 4.79, *p* < .001).

Finally, there was not a main effect of sign, showing that enriching by signs does not improve learning over no enriching. However, an interaction between concreteness and sign was also found (*Estimate* = -0.07, *SE* = 0.03, *z* = -2.26, *p* = .02), although post-hoc tests revealed no significant differences between specific conditions ($M_{abstract\ sign}$ = 0.28, $SD_{abstract\ sign}$ = 0.45, $M_{abstract\ no\ sign}$ = 0.25, $SD_{abstract\ no\ sign}$ = 0.43; *Estimate* = 0.09, *SE* = 0.08, *z* = 1.19, *p* = 0.23; $M_{concrete\ sign}$ = 0.33, $SD_{concrete\ sign}$ = 0.47, $M_{concrete\ no\ sign}$ = 0.35, $SD_{concrete\ no\ sign}$ = 0.48; *Estimate* = -0.02, *SE* = 0.06, *z* = -0.28, *p* = 0.78).

2.2.2. Iconic versus non-iconic sign

Only trials in which a sign was presented were included in this analysis. Independent variables were *test time* (T1, T2, T3, T4), *concreteness* (concrete versus abstract) and *iconicity* (iconic versus non-iconic sign).

There was again a main effect of test time (*Estimate* = 0.40, *SE* = 0.04, *z* = 10.62, *p* < .001). Post-hoc analyses showed higher accuracy in T2 versus T1 (M_{T1} = 0.07, SD_{T1} = 0.26, M_{T2} = 0.31, SD_{T2} = 0.46; *Estimate* = 1.07, *SE* = 0.10, *z* = 10.57, *p* < .001); higher accuracy in T3 versus T1 (M_{T3} = 0.51, SD_{T3} = 0.50; *Estimate* = 1.61, *SE* = 0.12, *z* = 13.70, *p* < .001); higher accuracy in T4 versus T1 (M_{T4} = 0.33, SD_{T4} = 0.47; *Estimate* = 1.08, *SE* = 0.11, *z* = 10.08, *p* <.001); higher accuracy in T3 versus T2 (*Estimate* = 0.56, *SE* = 0.05, *z* = 10.61, *p* < .001); no significant differences between T2 and T4 (*Estimate* = 0.04, *SE* = 0.05, *z* = -0.67, *p* = .50); and higher accuracy for T3 versus T4 (*Estimate* = -0.53, *SE* = 0.51, *z* = -10.20, *p* < .001).

We also observed a main effect of concreteness (*Estimate* = 0.18, *SE* = 0.07, *z* = 2.67, *p* = .01), with higher accuracy in the recall of concrete words (*M* = 0.33, *SD* = 0.47) than abstract words (*M* = 0.28, *SD* = 0.45).

The effect of concreteness interacted with test time (*Estimate* = 0.06, *SE* = 0.02, *z* = 2.40, *p* = .02). Post-hoc analyses revealed no differences in recall between concrete and abstract words at T1 ($M_{concrete}$ = 0.07, $SD_{concrete}$ = 0.26, $M_{abstract}$ = 0.08, $SD_{abstract}$ = 0.27; *Estimate* = 0.00, *SE* = 0.15, *z* = 0.01, *p* = .99) and T2 ($M_{concrete}$ = 0.33, $SD_{concrete}$ = 0.47, $M_{abstract}$ = 0.30, $SD_{abstract}$ = 0.46; *Estimate* = 0.13, *SE*

= 0.07, z = 1.81, p = .07), but concrete words were better recalled than abstract words at T3 ($M_{concrete}$ = 0.55, $SD_{concrete}$ = 0.50, $M_{abstract}$ = 0.47, $SD_{abstract}$ = 0.50; *Estimate* = 0.23, *SE* = 0.07, z = 3.31, p < .001) and T4 ($M_{concrete}$ = 0.37, $SD_{concrete}$ = 0.48, $M_{abstract}$ = 0.28, $SD_{abstract}$ = 0.45; *Estimate* = 0.29, *SE* = 0.08, z = 3.87, p < .001).

Finally, and crucial for our aims, the main effect of iconicity approached significance (*Estimate* = 0.11, *SE* = 0.06, z = 1.84, p = .06) and there was an interaction between concreteness and iconicity (*Estimate* = -0.11, *SE* = 0.04, z = -3.02, p = .01). Post-hoc analyses revealed higher recall for abstract words if the accompanying sign was iconic (M_{AI} = 0.32, SD_{AI} = 0.46, M_{ANI} = 0.25, SD_{ANI} = 0.43; *Estimate* = 0.24, *SE* = 0.10, z = 2.47, p = .01). In contrast, accuracy for concrete words did not vary as a function of sign iconicity (M_{CI} = 0.33, SD_{CI} = 0.47, M_{CNI} = 0.33, SD_{CNI} = 0.47; *Estimate* = -0.02, *SE* = 0.09, z = -0.21, p = .83) (see Figure 5).

2.3. Discussion

Study 1 explored the facilitating effect of iconic and non-iconic signs on abstract and concrete FVL by using signs from LSE without enactment. The main finding was the interaction between concreteness and iconicity: iconic signs helped participants to learn abstract words in a new language, whereas iconicity was not relevant for concrete words. Having in mind the theoretical principles of the CMT (Lakoff and Johnson, 1980; Lakoff et al., 1999) and their application to the iconicity of signs in sign languages (Taub, 2001; Wilcox, 2000), it can be argued that iconic signs help participants to find a concrete source domain where to ground abstract words, thereby improving their learning.

We also replicated the well-established concreteness effect, as words that refer to concrete concepts were more frequently recalled than words for abstract concepts (Kaushanskaya and Rechtzigel, 2012). We did not replicate the enrichment effect, but the concreteness effect interacted with sign enrichment, in the direction of abstract concepts showing greater numeric benefits in label learning than concrete concepts, although the effect was not strong enough to surface in post-hoc comparisons. Concrete concepts are richer in sensorimotor information (Barsalou et al., 2003; Macedonia and Knösche, 2011), what makes easier to learn new labels for them. We suggest that this makes their label learning also less dependent than abstract concepts on helping factors such as enrichment.

There was also a main effect of test time which showed that participants progressively increased their recall for newly learned words during the first session (from T1 to T2, and from T2 to T3), and that recall decayed but was still detectable one week later. This finding shows that our methodology, in which participants are only trained for 45 minutes (15 minutes per block) delivers fast and lasting results. In the previous literature using gesture enrichment, the only other study in which participants learnt for just one day is Repetto et al. (2017), which did not find a facilitating effect of gestures on learning.

In study 1 participants were instructed to repeat the Vimmi words aloud in the repetition phase of the trial, but they did not have to enact the sign. Given the lack of consensus regarding the importance of enactment for FVL, we conducted study 2. Study 2 was identical to study 1, but participants were asked to both repeat the word aloud and enact the sign in the repetition phase.

3. Study 2

3.1. Method

3.1.1. Participants

Forty-one students from the University of Granada (Spain; mean age = 21.46, *SD* = 3.56, seven males, three left-handed) took part in the study, and were rewarded with course credits. All participants were Spanish native speakers, had no prior knowledge of any sign language and did not participate in study 1. All had normal or corrected-to-normal vision and no hearing impairments. The same informed consent and ethical guidelines were followed as in study 1. The study was also approved by the ethics committee of Edge Hill University's Department of Psychology.

3.1.2. Materials and procedure

The same materials as in study 1 were used. The procedure was as in study 1, with one critical difference: On the "repetition" phase of the trial, participants were asked to both repeat the L2 word aloud and to enact the sign as it appeared on the screen on sign trials.

3.2. Results

Data were analysed as in study 1, using the same models. Full disclosure of fixed and random effects, together with estimates, standard errors, z scores and p values for the slopes for all the analyses reported here can be found in Tables 3 and 4 at the OSF project (see below).

3.2.1. Sign versus no sign

Independent variables were *test time* (T1, T2, T3, T4), *concreteness* (concrete versus abstract) and *sign* (sign versus no sign).

There was a main effect of test time (*Estimate* = 0.63, *SE* = 0.04, z = 14.13, $p < .001$). Post-hoc analyses showed higher accuracy in T2 versus T1 (M_{T1} = 0.06, SD_{T1} = 0.23, M_{T2} = 0.27, SD_{T2} = 0.45; *Estimate* = 0.97, *SE* = 0.08, z = 12.75, $p < .001$); higher accuracy in T3 versus T1 (M_{T3} = 0.46, SD_{T3} = 0.50; *Estimate* = 1.46, *SE* = 0.09, z = 16.10, $p < .001$); higher accuracy in T4 versus T1 (M_{T4} = 0.31, SD_{T4} = 0.46; *Estimate* = 1.04, *SE* = 0.10, z = 10.00, $p < .001$); higher accuracy in T3 versus T2 (*Estimate* = 0.50, *SE* = 0.05, z = 10.99, $p < .001$); no significant differences between T2 and T4 (*Estimate* = 0.07, *SE* = 0.06, z = 1.06, p = .29); and higher accuracy for T3 versus T4 (*Estimate* = -0.40, *SE* = 0.11, z = -3.53, $p < .001$).

We also observed a main effect of concreteness (*Estimate* = 0.22, *SE* = 0.07, z = 2.97, p = .01), with higher accuracy in the recall of labels of concrete (*M* = 0.31, *SD* = 0.46) than abstract concepts (*M* = 0.24, *SD* = 0.43). The effect of concreteness did not interact with test time. There was no main effect of sign enrichment nor interaction with concreteness.

3.2.2. Iconic versus non-iconic signs

This analysis was carried out only on trials where a sign was presented. Independent variables were *test time* (T1, T2, T3, T4), *concreteness* (concrete versus abstract) and *iconicity* (iconic versus non-iconic sign). There was a main effect of test time (*Estimate* = 0.72, *SE* = 0.05, z = 13.94, $p < .001$). Post-hoc analyses showed higher accuracy in T2 versus T1 (M_{T1} = 0.06, SD_{T1} = 0.23, M_{T2} = 0.27, SD_{T2} = 0.44; *Estimate* = 1.01, *SE* = 0.09, z = 10.86, $p < .001$); higher accuracy in T3 versus T1 (M_{T3} = 0.47, SD_{T3} = 0.50; *Estimate* = 1.56, *SE* = 0.11, z = 14.14, $p < .001$); higher accuracy in T4 versus T1 (M_{T4} = 0.32, SD_{T4} = 0.47; *Estimate* = 1.13, *SE* = 0.12, z = 9.51, $p < .001$); higher accuracy in T3 versus T2 (*Estimate* = 0.57, *SE* = 0.05, z = 10.83, $p < .001$); no significant differences between T2 and T4 (*Estimate* = 0.16, *SE* = 0.10, z = 1.61, p = .11); and higher accuracy for T3 versus T4 (*Estimate* = -0.41, *SE* = 0.11, z = -3.73, $p < .001$).

We observed a main effect of concreteness (*Estimate* = -0.27, *SE* = 0.11, z = 2.5, p = .01), with higher accuracy in the recall of concrete words (*M* = 0.31, *SD* = 0.46) than abstract words (*M* = 0.25, *SD* = 0.43). There was also a main effect of iconicity (*Estimate* = 0.17, *SE* = 0.08, z = 2.12, p = .03): words paired with iconic signs were more accurately recalled (*M* = 0.31, *SD* = 0.46) than words paired with non-iconic signs (*M* = 0.25, *SD* = 0.44). Finally, the interaction between concreteness and iconicity was also replicated (*Estimate* = -0.13, *SE* = 0.04, *z* = -3.10, $p < .01$). Post-hoc contrasts revealed higher recall of labels for abstract concepts when the accompanying signs were iconic (M_{AI} = 0.29, SD_{AI} = 0.45, M_{ANI} = 0.20, SD_{ANI} = 0.40; *Estimate* = 0.30, *SE* = 0.10, *z* = 3.05, $p < 0.01$), while labels of concrete concepts were not helped by the iconicity of the sign (M_{CI} = 0.33, SD_{CI} = 0.47, M_{CNI} = 0.30, SD_{CNI} = 0.46; *Estimate* = 0.07, *SE* = 0.08, *z* = 0.83, *p* = .41; see Figure 5).

3.3. Discussion

Study 2 was identical to study 1 in all respects but one: the participants enacted the signs that accompanied the new labels to be learnt. Overall, enrichment by enacted iconic signs did not benefit FVL over a no-enactment condition, as in study 1, nor did it interact with concreteness. We postpone the discussion of the enactment effect to the direct contrast between studies 1 and 2 that we report in the next section. All other main findings of study 1 were replicated in study 2: There was a main effect of iconicity, with higher recall of words paired with iconic (versus non-iconic) signs, and iconicity helped label learning for abstract concepts but did not significantly impact on label learning for concrete concepts. In addition, the main effects of concreteness and test time observed in study 1 were replicated in study 2, although in this case the learning curve was not different for labels of concrete and abstract concepts.

4. Comparison of studies 1 and 2

To assess whether there is an enactment effect, a joint analysis including studies 1 and 2 as a between-participants factor was performed. The joint analysis followed the guidelines of the prior analyses and used similar models. Full disclosure of fixed and random effects, together with estimates, standard errors, z scores and p values for the slopes for all the analyses reported here can be found in Tables 5 and 6 at the OSF project (see below).

4.1. Sign versus no sign

Independent variables were *study* (study 1 versus study 2), *test time* (T1, T2, T3, T4), *concreteness* (concrete versus abstract), and *sign* (sign versus no sign). There was a main effect of study (*Estimate* = -0.47, *SE* = 0.06, *z* = -7.98, $p < .001$), showing that participants learned more if they did not enact the sign ($M_{study\ 1}$ = 0.30, SD_{study1} = 0.46; M_{study2} = 0.27, SD_{study2} = 0.45).

There was also a main effect of test time (*Estimate* = 0.56, *SE* = 0.03, *z* = 18.56, *p* < .001) with the same learning curve than in studies 1 and 2: higher accuracy in T2 versus T1 (M_{T1} = 0.07, SD_{T1} = 0.25, M_{T2} = 0.29, SD_{T2} = 0.46; *Estimate* = 0.98, *SE* = 0.05, *z* = 19.21, *p* < .001); higher accuracy in T3 versus T2 (M_{T3} = 0.48, SD_{T3} = 0.50 *Estimate* = 0.51, *SE* = 0.03, *z* = 15.65, *p* < .001); and lower accuracy for T4 versus T3 (M_{T4} = 0.31, SD_{T4} = 0.46 *Estimate* = -0.43, *SE* = 0.05, *z* = -8.16, *p* < .001), although still better recall than at T1 (*Estimate* = 1.04, *SE* = 0.07, *z* = 15.75, *p* < .001).

We observed a main effect of concreteness (*Estimate* = -0.22, *SE* = 0.05, *z* = 4.33, *p* < .001) with higher accuracy in the recall of concrete words (*M* = 0.32, *SD* = 0.46) than abstract words (*M* = 0.26, *SD* = 0.44). There was no main effect of sign enrichment.

The four fixed factors entered in interactions with some other factors. The study factor interacted with test time (*Estimate* = 0.11, *SE* = 0.22, *z* = 4.91, *p* < .001), but post-hoc analyses showed no differences between study 1 and 2 in any of the evaluation times (T1 (M_{study1} = 0.07, SD_{study1} = 0.26, M_{study2} = 0.06, SD_{study2} = 0.23; *Estimate* = -0.10, *SE* = 0.13, *z* = -0.77, *p* = .44), T2 (M_{study1} = 0.31, SD_{study1} = 0.46, M_{study2} = 0.27; SD_{study2} = 0.45; *Estimate* = -0.12, *SE* = 0.11, *z* = -1.11, *p* = 0.27), T3 (M_{study1} = 0.51; SD_{study1} = 0.5, M_{study2} = 0.46; SD_{study2} = 0.50; *Estimate* = -0.14, *SE* = 0.13, *z* = -1.05, *p* = .29) or T4 (M_{study1} = 0.32, SD_{study1} = 0.46, M_{study2} = 0.31, SD_{study2} = 0.46; *Estimate* = 0.05, *SE* = 0.04, *z* = 1.28, *p* = .2)).

Concreteness interacted with test time (*Estimate* = 0.05, *SE* = 0.02, *z* = 3.30, *p* < .001). Post-hoc analyses showed better label learning for concrete over abstract concepts at all times with greater numerical differences as learning progressed (T1: $M_{concrete}$ = 0.07, $SD_{concrete}$ = 0.26, $M_{abstract}$ = 0.06, $SD_{abstract}$ = 0.24, *Estimate* = 0.24, *SE* = 0.09, *z* = 2.77, *p* < 0.01; T2: $M_{concrete}$ = 0.32, $SD_{concrete}$ = 0.47, $M_{abstract}$ = 0.27, $SD_{abstract}$ = 0.44, *Estimate* = 0.19, *SE* = 0.04, *z* = 4.64, *p* < .001; T3: $M_{concrete}$ = 0.53, $SD_{concrete}$ = 0.50, $M_{abstract}$ = 0.44, $SD_{abstract}$ = 0.50, *Estimate* = 0.25, *SE* = 0.04, *z* = 5.94, *p* < .001; T4: $M_{concrete}$ = 0.37, $SD_{concrete}$ = 0.48, $M_{abstract}$ = 0.26, $SD_{abstract}$ = 0.44, *Estimate* = 0.30, *SE* = 0.04, *z* = 7.51, *p* < .001).

Finally, there was a significant interaction between test time and sign (*Estimate* = 0.03, *SE* = 0.02, *z* = 1.97, *p* = .05). Nevertheless, post-hoc analyses showed no significant differences between sign and no sign at any test time (T1 M_{sign} = 0.06, SD_{sign} = 0.25, $M_{no\ sign}$ = 0.07, $SD_{no\ sign}$ = 0.25, *Estimate* = -0.08, *SE* = 0.10, *z* = -0.79, *p* = .43; T2 M_{sign} = 0.29, SD_{sign} = 0.45, $M_{no\ sign}$ = 0.30, $SD_{no\ sign}$ = 0.46, *Estimate* = -0.06, *SE* = 0.05, *z* = -1.05, *p* = 0.29; T3 M_{sign} = 0.49, SD_{sign} = 0.50, $M_{no\ sign}$ = 0.46, $SD_{no\ sign}$ = 0.50, *Estimate* = 0.77, *SE* = 0.04, *z* = 1.70 *p* = .08; T4 M_{sign} = 0.32, SD_{sign} = 0.47, $M_{no\ sign}$ = 0.29, *SD* = 0.45, *Estimate* = 0.09, *SE* = 0.05, *z* = 1.83, *p* = .07).

4.2. Iconic versus non-iconic sign

Independent variables were *study* (study 1 versus study 2), *test time* (T1, T2, T3, T4), *concreteness* (concrete versus abstract), and *iconicity* (iconic versus non-iconic sign). The analysis on the subset of sign trials found a main effect of study (*Estimate* = -0.48, *SE* = 0.10, z = -0.79, p = 0.43), with participants being more accurate in study 1 without enactment (M = 0.31, SD = 0.46) than in study 2 with enactment (M = 0.28, SD = 0.45).

Moreover, we observed a main effect of test time (*Estimate* = -0.62, *SE* = 0.03, z = 18.36, $p < .001$) with the same pattern than in studies 1 and 2: accuracy grew from T1 to T2 (M_{T1} = 0.06, SD_{T1} = 0.25, M_{T2} = 0.29, SD_{T2} = 0.45; *Estimate* = 0.99, *SE* = 0.65, z = 15.29, $p < .001$); and did again from T2 to T3 (M_{T3} = 0.49, SD_{T3} = 0.50; *Estimate* = 0.56, *SE* = 0.04, z = 15.09, $p < .001$); and decreased from T3 to T4 (M_{T4} = 0.32, SD_{T4} = 0.47; *Estimate* = -0.41, *SE* = 0.05, z = -8.02, $p < .001$), although T4 still had greater recall than T1 (*Estimate* = 1.11, *SE* = 0.08, z = 14.34, $p < .001$).

We also observed a main effect of concreteness (*Estimate* = 0.20, *SE* = 0.07, z = 3.07, $p < .01$) with better recall of labels for concrete (M = 0.32, SD = 0.47) than abstract concepts (M = 0.26, SD = 0.44).

Crucially, there was also a main effect of iconicity (*Estimate* = 0.15, *SE* = 0.05, z = 3.05, $p < .01$), with words paired with iconic signs being more accurately recalled (M = 0.31, SD = 0.46) than words paired with non-iconic signs (M = 0.27, SD = 0.44).

Study and test time interacted (*Estimate* = 0.13, *SE* = 0.02, z = 5.28, $p < .001$), but post-hoc analyses did not find differences between the studies at any test time (M_{study1} = 0.07, SD_{study1} = 0.26, M_{study2} = 0.06, SD_{study2} = 0.23; *Estimate* = -0.15, *SE* = 0.16, z = -0.94, p = .35), T2 (M_{study1} = 0.31, SD_{study1} = 0.46, M_{study2} = 0.27; SD_{study2} = 0.44; *Estimate* = -0.13, *SE* = 0.11, z = -1.18, p = .24), T3 (M_{study1} = 0.51; SD_{study1} = 0.50, M_{study2} = 0.47; SD_{study2} = 0.50; *Estimate* = -0.10, *SE* = 0.13, z= -0.73, p =.46) or T4 (M_{study1} = 0.33, SD_{study1} = 0.47, M_{study2} = 0.32, SD_{study2} = 0.47; *Estimate* = 0.05, *SE* = 0.05, z = 1.03, p = .30).

Importantly, there was an interaction between concreteness and iconicity (*Estimate* = -0.12, *SE* = 0.03, z = -4.26, $p < .01$), revealing better label learning for abstract concepts when the accompanying signs were iconic (M_{AI} = 0.30, SD_{AI} = 0.46, M_{ANI} = 0.23, SD_{ANI} = 0.42; *Estimate* = 0.24, *SE* = 0.07, z = 3.51, $p < .001$), whereas label learning for concrete concepts did not benefit from the iconicity of the sign (M_{CI} = 0.33, SD_{CI} = 0.47, M_{CNI} = 0.32, SD_{CNI} = 0.46; *Estimate* = 0.02, *SE* = 0.07, z = 0.30, p = .76).

4.3. Comparison of iconic sign versus no sign and non-iconic sign versus no sign

To further explore the particular benefit that might be offered by iconic signs, and taking advantage of the greater statistical power afforded by the joint analysis of studies 1 and 2, we ran additional analyses comparing recall for Vimmi words paired with iconic signs versus without signs on one hand, and for those paired with non-iconic signs versus without signs on the other hand. Means, SDs and post-hoc analyses are reported only for key comparisons and interactions that offer new information (i.e., those involving the variable "stimuli"). Full disclosure of fixed and random effects, together with estimates, standard errors, z scores and p values for the slopes for all the analyses reported here can be found in Tables 7 and 8 at the OSF project (see below).

In the first partial design, independent variables were *study* (study 1 versus study 2), *test time* (T1, T2, T3, T4), *concreteness* (concrete versus abstract), and *stimuli* (iconic versus no sign). There was a main effect of test time (*Estimate* = 0.54, *SE* = 0.04, z = 14.47, $p < .001$), and an interaction between test time and concreteness (*Estimate* = 0.07, *SE* = 0.02, z = 3.02, p = .002). Importantly, there was an interaction between test time and stimuli (*Estimate* = 0.05, *SE* = 0.02, z = 2.46, p = .01). Post hoc analyses found no differences between iconic signs and no sign at T1 ($M_{iconic\ sign}$ = 0.07, $SD_{iconic\ sign}$ = 0.06, $M_{no\ sign}$ = 0.07, $SD_{no\ sign}$ = 0.06; *Estimate* = -0.08, *SE* = 0.12, z = -0.71, p = .48) or T2 ($M_{iconic\ sign}$ = 0.32, $SD_{iconic\ sign}$ = 0.16, $M_{no\ sign}$ = 0.30, $SD_{no\ sign}$ = 0.14; *Estimate* = 0.02, *SE* = 0.05, z = 0.43, p = .67); however, words paired with iconic signs were better remembered than those paired with no signs at T3 ($M_{iconic\ sign}$ = 0.52, $SD_{iconic\ sign}$ = 0.20, $M_{no\ sign}$ = 0.46, $SD_{no\ sign}$ = 0.20; *Estimate* = 0.15, *SE* = 0.05, z = 2.99, p = .002) and T4 ($M_{iconic\ sign}$ = 0.35, *SD* = 0.15, $M_{no\ sign}$ = 0.29, $SD_{no\ sign}$ = 0.16; *Estimate* = 0.17, *SE* = 0.06, z = 2.81, p = .005). Thus, when enrichment by iconic signs is compared to no enrichment at all, learning benefits are detected that grow larger as learning progresses and remain after a retention period of one week.

In the second partial design, independent variables were *study* (study 1 versus study 2), *test time* (T1, T2, T3, T4), *concreteness* (concrete versus abstract), and *stimuli* (non-iconic versus no sign). There was a main effect of test time (*Estimate* = 0.50, *SE* = 0.04, z = 12.59, $p < .001$) and an interaction between test time and concreteness (*Estimate* = 0.09, *SE* = 0.02, z = 3.73, $p < .001$). Thus, when enrichment by non-iconic signs is compared to no enrichment at all, no learning benefits are observed at any of the testing times.

5. General discussion

We explored the facilitating effect of iconic and non-iconic signs on the learning of new labels for abstract and concrete concepts by using signs from Spanish Sign Language (LSE), which could be enacted (study 2) or not (study 1). Our main research question was whether iconic signs from sign languages can be an effective tool to teach abstract foreign vocabulary. Based on the Conceptual Metaphor Theory (Lakoff and Johnson, 1980; Lakoff et al., 1999) and its application to sign languages (Taub, 2001; Wilcox, 2000), and on prior experimental work on gestures and FVL (e.g., Macedonia and Knösche, 2011; see review in the Introduction), we expected this to be the case. Present data supported this hypothesis.

Overall, we found a main effect of iconicity: Vimmi words paired with iconic signs were more accurately recalled than those paired with non-iconic signs. This goes in line with some previous literature using iconic versus non-iconic gestures (Garcia-Gamez and Macizo, 2019; Kelly et al., 2009). This effect in our study, however, interacts with concreteness, such that iconicity facilitates abstract, but not concrete FVL. Concrete words are richer in sensorimotor information than abstract words (Barsalou et al., 2003; Macedonia and Knösche, 2011), and, in our studies, concrete words were easier to learn than abstract words. In this context, it seems intuitive that the iconicity of the accompanying sign exerts a stronger effect on the more difficult to learn labels, those for abstract concepts. We suggest that the iconic signs help the learner to ground the labels for abstract concepts onto a concrete source (Taub, 2001). This pattern of results is generally compatible with the findings by Morett (2015) regarding the learning of the signs themselves by hearing non-signers when the learning is tested after a retention period of at least one week. To the best of our knowledge, this is the first report of an interaction between concreteness and iconicity of the accompanying stimulus in FVL.

We had three additional and less central research questions. First, whether iconic signs can facilitate FVL for concrete concepts over non-iconic signs, as expected from some previous literature (e.g., Garcia-Gamez and Macizo, 2019; Kelly et al., 2009). As described above, present results suggest that the effect of enrichment by iconic signs (over non-iconic signs) is limited to the learning of words for abstract concepts.

The second additional research question was whether the mere observation of an accompanying sign (that is, any sign without enactment) can enhance FVL. Some authors (as Huang et al., 2019) have argued that any gesture can enhance learning if the learner attaches meaning to it. This goes in line with the Dual Coding Theory (Clark and Paivio, 1991), as engaging more than one perceptual channel should lead to facilitation. Again, present results do not support this claim, as we did not observe neither a general effect of sign enrichment, nor a benefit of presenting non-iconic signs versus no signs. The passive viewing of non-iconic signs (versus a static image) was not enough to facilitate FVL. Rather, our data show that sign iconicity is helpful for FVL, and particularly so for abstract words.

Our final research question explored whether enactment produces extra learning benefits in FVL. For Craik and Lockhart (1972), the additional enrichment offered by enactment would lead to a deeper processing level, potentially providing more grounding for abstract words, and, thus, more recall. Following the embodiment theory (Barsalou et al., 2003; Meteyard et al., 2012), enactment further enriches the stimuli by involving the body and should lead to better learning of new labels. Contrary to these expectations, present results showed that participants recalled more words without enactment (study 1) than with it (study 2). As discussed in the introduction, very few studies have carried out direct comparisons between a condition with and another without enactment, and the result that is more often reported is the lack of significant differences (de Nooijer et al., 2014; Krönke et al., 2013; Macedonia et al., 2014). As far as we can tell, the present study is also the first to report a detrimental effect of enactment in this literature. This effect agrees more with positions that emphasise that increasing cognitive load during learning can hamper recall (e.g., Mayer and Moreno, 2003), as having to enact the gesture besides watching it may increase cognitive load. It is worth noting that the present studies may have also been more cognitively demanding than some previous work because conditions were mixed in the same block, while previous studies present them in different blocks (e.g., Garcia-Gamez & Macizo, 2019; Huang et al., 2019; Kelly, 2009; Macedonia and Knösche, 2011; Mayer et al., 2017; Morett, 2019; Repetto et al. 2017), precisely to avoid increasing cognitive load (Gamez and Macizo, 2019). Thus, it is possible that enacting the gesture only hampers FVL when the participant is already under high cognitive load. Finding out will need more research directly comparing high and low cognitive load conditions.

All in all, the present study makes both theoretical and applied contributions. In the theoretical front, it supports the main tenet of the Conceptual Metaphor Theory: that the representation of abstract concepts is mediated by metonymies and conceptual metaphors that link those ill-understood concepts with concrete concepts of which we have a better understanding given their richer sensory-motor information (Lakoff and Johnson, 1980; Lakoff et al., 1999); or, in simpler words, that iconic signs help participants to find a concrete source where to ground abstract meaning. This rationale also agrees with the embodiment theory (Meteyard et al., 2012) for which higher order processes (such as language) are rooted in perception and action. Using metonymies and metaphors allows people to root those concepts that do not have clear sensory-motor features (i.e., they are abstract) onto others that do have them (i.e., they are concrete) providing them with a "physical appearance". Present data, however, do not support strong views of embodiment for which semantics completely depend on the body (e.g., Zwaan and Ross, 2004; see Meteyard et al., 2012). If that were the case, we should have observed a beneficial effect of enactment. Rather, we believe that sensory-motor systems mediate in higher order processes (e.g., Vigliocco et al., 2004).

From an applied viewpoint, our study could open the door to improvements in foreign language teaching, specially of vocabulary that refers to abstract concepts. Sign languages offer to the FVL teacher a rich vocabulary that refers to abstract concepts. Coupled with linguistics analyses that are able to discriminate between those signs that have an iconic relation to the abstract concepts from those that do not, the teacher has a rich and ready-made source of signs to assist the learning of these especially difficult words. The same advantage is offered to the FVL researcher: from now on it will not be necessary to resort to subjective intuition when selecting the gestures to enrich word learning. Even when the interest is in the use of non-conventional gestures, the same analysis based on metonymies and conceptual metaphors has also been applied to gestures (Cienki and Müller, 2008) and it can offer a principled guidance to their generation and selection.

Many questions remain unanswered. To mention just a couple of examples: What is exactly the right amount of training that can generate effective word learning and that can benefit from enrichment by signs? Most previous studies trained participants in a small set of words for days. Repetto et al. (2017) provided a single session of 35 minutes and did not find any gesture enrichment effect, what made them suggest that the facilitatory effect of gestures for FVL may need extended practice. Present data show, however, that iconic signs were able to improve the learning of new labels for abstract concepts after three presentations in a single session. Further research is needed to examine the effects of different exposure conditions on the strength and duration of the effects. Another pending question is whether signs will also help the learning of other word classes such as verbs and adjectives. We only assessed nouns to eliminate potential confounds and to maximise word learning (Macedonia and Knösche, 2011; Marshall, 2003; Waxman et al., 2013). Previous literature has shown promising results using different word classes and experimenter-created gestures instead of signs (Macedonia and Knösche, 2011; Garcia-Gamez and Macizo, 2019). Examining word class and FVL with signs in the same study will hopefully broaden our understanding of the mechanisms behind the influence of sign iconicity on FVL.

6. Conclusions

Hereby, for the first time we demonstrate that iconicity in the vocabulary of sign languages can be effectively used to teach abstract vocabulary in hearing non-signers learning a foreign oral language. This result significantly contributes to support the tenets of the Conceptual Metaphor and embodiment theories. It also has relevant implications for language learning methodologies and research practices, which may use the vocabulary of sign languages to improve the selection of enrichment materials and, therefore, achieving better learning.

The OSF project of this article can be found at: https://osf.io/9axke/?view_only=cabc9bdc9aae4444ae3451119a454e7b

Acknowledgements:

The author(s) would like to thank all the participants in this research.

Declaration of Conflicting Interest:

The author(s) declare(s) that there is no conflict of interest.

Funding:

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Edge Hill University's Research Investment Fund [grant number 1RODRI16].

7. References

Andrä C, Mathias B, Schwager A, Macedonia M and 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*: 1-36.

Barsalou LW, Simmons WK, Barbey, AK, and Wilson CD (2003) Grounding conceptual knowledge in modality-specific systems. *Trends in cognitive sciences* 7(2): 84-91.

Bates D (2010) lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-33.

Bergmann K and Macedonia M (2013) A virtual agent as vocabulary trainer: iconic gestures help to improve learners' memory performance. In: *International workshop on intelligent virtual agents*, Edinburgh, UK, 29-31 August 2013, pp. 139-148. Springer, Berlin, Heidelberg.

Borghi AM and Setti A (2017) Abstract concepts and aging: An embodied and grounded perspective. *Frontiers in Psychology* 8, 430.

Blasi DE, Wichmann S, Hammarström H, Stadler PF and Christiansen MH (2016) Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences* 113(39), 10818-10823.

Caselli NK and Pyers JE (2017) The road to language learning is not entirely iconic: Iconicity, neighborhood density, and frequency facilitate acquisition of sign language. *Psychological science 28*(7) 979-987.

Cienki AJ and Müller C (2008) *Metaphor and gesture*. John Benjamins Pub. Co.

Clark JM and Paivio A (1991) Dual coding theory and education. *Educational psychology review* 3(3): 149-210.

Craik FI and Lockhart RS (1972) Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior* 11(6): 671-684.

de Nooijer JA, van Gog T, Paas F and Zwaan RA (2014) Words in action: using gestures to improve verb learning in primary school children. *Gesture 14*(1): 46-69.

Dingemanse M, Blasi DE, Lupyan G, Christiansen MH and Monaghan, P (2015) Arbitrariness, iconicity and systematicity in language. *Trends in Cognitive Sciences* 19(10): 603–615.

Dove G (2015) How to go beyond the body: an introduction. *Frontiers in Psychology* 6 660.

Duchon A, Perea M, Sebastián-Gallés N, Martí A and Carreiras M (2013) EsPal: one-stop shopping for Spanish word properties. *Behavior Research Methods* 45: 1246-1258.

Ettlinger M, Morgan-Short K, Faretta-Stutenberg M and Wong PC (2016) The relationship between artificial and second language learning. *Cognitive science* 40(4): 822-847.

Frishberg N (1975) Arbitrariness and iconicity: historical change in American Sign Language. *Language*: 696-719.

Garcia-Gamez AB and Macizo P (2019) Learning nouns and verbs in a foreign language: The role of gestures. *Applied Psycholinguistics* 40(2): 473-507.

Guasch M, Boada R, Ferré P and Sánchez-Casas R (2013) NIM: A Web-based Swiss Army knife to select stimuli for psycholinguistic studies. *Behavior Research Methods* 45: 765-771.

Guasch M, Ferré P and Fraga I (2016) Spanish norms for affective and lexico-semantic variables for 1,400 words. *Behavior Research Methods* 48: 1358-1369.

Hinojosa JA, Haro J, Magallares S, Duñabeitia JA and Ferré P (2020) Iconicity ratings for 10,995 Spanish words and their relationship with psycholinguistic variables. *Behavior Research Methods* 1-14.

Huang X, Kim N and Christianson K (2019) Gesture and vocabulary learning in a second language. *Language Learning* 69(1): 177-197.

Imai M and Kita S (2014) The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. Philosophical transactions of the Royal Society B: *Biological sciences* 369(1651) 20130298.

Kaushanskaya M and Rechtzigel K (2012) Concreteness effects in bilingual and monolingual word learning. *Psychonomic bulletin & review* 19(5): 935-941.

Kelly SD, McDevitt T and 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.

Knoeferle K, Li J, Maggioni E and Spence C (2017) What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. *Scientific reports* 7(1): 1-11.

Kroenke KM, Kraft I, Regenbrecht and Obrig H (2013) Lexical learning in mild aphasia: Gesture benefit depends on patholinguistic profile and lesion pattern. *Cortex* 49(10): 2637-2649.

Krönke KM, Mueller K, Friederici AD and Obrig H (2013) Learning by doing? The effect of gestures on implicit retrieval of newly acquired words. *Cortex* 49(9): 2553-2568.

Lakoff G and Johnson M (1980) Conceptual metaphor in everyday language. *The journal of Philosophy* 77(8): 453-486.

Lakoff G, Johnson M and Sowa JF (1999) Review of Philosophy in the Flesh: The embodied mind and its challenge to Western thought. *Computational Linguistics* 25(4): 631-634.

Marshall J (2003) Noun–verb dissociations—evidence from acquisition and developmental and acquired impairments. *Journal of Neurolinguistics* 16(2-3): 67-84.

Macedonia M (2014) Bringing back the body into the mind: gestures enhance word learning in foreign language. *Frontiers in psychology* 5 1467: 1-6.

Macedonia M, Bergmann K and Roithmayr F (2014) Imitation of a pedagogical agent's gestures enhances memory for words in second language. *Science Journal of Education* 2(5): 162-169.

Macedonia M and Knösche TR (2011) Body in mind: How gestures empower foreign language learning. *Mind, Brain, and Education* 5(4): 196-211.

Macedonia M and Mueller K (2016) Exploring the neural representation of novel words learned through enactment in a word recognition task. *Frontiers in psychology 7* 953: 1-14.

Macedonia M, Müller K and Friederici AD (2010) Neural correlates of high performance in foreign language vocabulary learning. *Mind, Brain, and Education* 4(3): 125-134.

Macedonia M, Müller K and Friederici AD (2011) The impact of iconic gestures on foreign language word learning and its neural substrate. *Human brain mapping* 32(6): 982-998.

Macedonia M, Repetto C, Ischebeck A and Mueller K (2019) Depth of encoding through observed gestures in foreign language word learning. *Frontiers in psychology* 10 33: 1-15.

Massaro DW and Perlman M (2017) Quantifying iconicity's contribution during language acquisition: Implications for vocabulary learning. *Frontiers in Communication 2*, 4.

Mayer RE and Moreno R (2003) Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist* 38(1): 43-52.

Mayer KM, Macedonia M and von Kriegstein K (2017) Recently learned foreign abstract and concrete nouns are represented in distinct cortical networks similar to the native language. *Human brain mapping* 38(9): 4398-4412.

Mayer KM, Yildiz IB, Macedonia M and von Kriegstein K (2015) Visual and motor cortices differentially support the translation of foreign language words. *Current Biology* 25(4): 530-535.

Meteyard L, Rodriguez-Cuadrado S, Bahrami B and Vigliocco G (2012) Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex* 48(7): 788-804.

Miozzo M, Villabol M, Navarrete E and Peressotti, F (2020) Hands show where things are: The close similarity between sign and natural space. *Cognition* 196 104106.

Morett LM (2015) Lending a hand to signed language acquisition: enactment and iconicity enhance sign recall in hearing adult American Sign Language learners. *Journal of Cognitive Psychology* 27(3): 251-276.

Morett LM (2018) In hand and in mind: Effects of gesture production and viewing on second language word learning. *Applied Psycholinguistics* 39(2): 355-381.

Morett LM (2019) The power of an image: Images, not glosses, enhance learning of concrete L2 words in beginning learners. *Journal of psycholinguistic research* 48(3): 643-664.

Nielsen AK and Dingemanse M (2020) Iconicity in word learning and beyond: A critical review. *Language and Speech*, 0023830920914339: 1-21.

Ortega G (2017) Iconicity and sign lexical acquisition: a review. *Frontiers in Psychology* 8 1280.

Perniss P and Vigliocco G (2014) The bridge of iconicity: from a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1651), 20130300: 1-13

Paivio A, Walsh M and Bons, T (1994) Concreteness effects on memory: When and why?. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20(5) 1196.

Perry LK, Perlman M and Lupyan G (2015) Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PloS one* 10(9) e0137147.

Porter A (2016) A helping hand with language learning: teaching French vocabulary with gesture. *The Language Learning Journal* 44(2): 236-256.

Repetto C, Pedroli E and Macedonia M (2017) Enrichment effects of gestures and pictures on abstract words in a second language. *Frontiers in psychology* 8 2136: 1-11.

Rowe ML, Silverman RD and Mullan BE (2013) The role of pictures and gestures as nonverbal aids in preschoolers' word learning in a novel language. *Contemporary Educational Psychology* 38(2): 109-117.

Schneider W, Eschman, A and Zuccolotto A (2007) E-Prime 2.0 professional. Pittsburgh, PA: Psychology Software Tools.

Sidhu DM, Williamson J, Slavova V and Pexman PM (2022) An investigation of iconic language development in four datasets. *Journal of Child Language* 49(2) 382-396.

Sutton-Spence R. and Woll B (1999) The linguistics of British Sign Language: an introduction. Cambridge University Press.

Taub SF (2000) Iconicity in American sign language: concrete and metaphorical applications. *Spatial Cognition and Computation* 2(1): 31-50.

Tellier M (2008) The effect of gestures on second language memorisation by young children. *Gesture* 8(2): 219-235.

Thompson RL, Vinson DP, Woll B and Vigliocco G (2012) The road to language learning is iconic: Evidence from British Sign Language. *Psychological science* 23(12): 1443-1448.

Van Hell JG and Mahn AC (1997) Keyword mnemonics versus rote rehearsal: Learning concrete and abstract foreign words by experienced and inexperienced learners. Language learning 47(3): 507-546.

Van Herreweghe M, De Meulder M and Vermeerbergen M (2015) From erasure to recognition (and back again?): The case of Flemish Sign Language. *The Oxford Handbook of Deaf Studies in Language Research, Policy and Practice*: 45-61.

Vigliocco G, Vinson DP, Lewis W and Garrett M F (2004) Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive psychology* 48(4), 422-488.

Vinson DP, Cormier K, Denmark T, Schembri A and Vigliocco G (2008) The British Sign Language (BSL) norms for age of acquisition, familiarity, and iconicity. *Behavior research methods 40*(4), 1079-1087.

Waxman S, Fu X, Arunachalam S, Leddon E, Geraghty K and Song HJ (2013) Are nouns learned before verbs? Infants provide insight into a long-standing debate. *Child development perspectives* 7(3): 155-159.

Wilcox PP (2000) *Metaphor in American sign language*. Gallaudet University Press.

Zwaan RA and Ross BH (2004) The immersed experiencer: Toward an embodied theory of language comprehension. *Psychology of learning and motivation* 44: 35-62.

Figure 1. The “gorilla” sign in Spanish Sign Language (LSE).

Figure 2. The “belief” sign in Spanish Sign Language (LSE).

Figure 3. Static image of the signer.

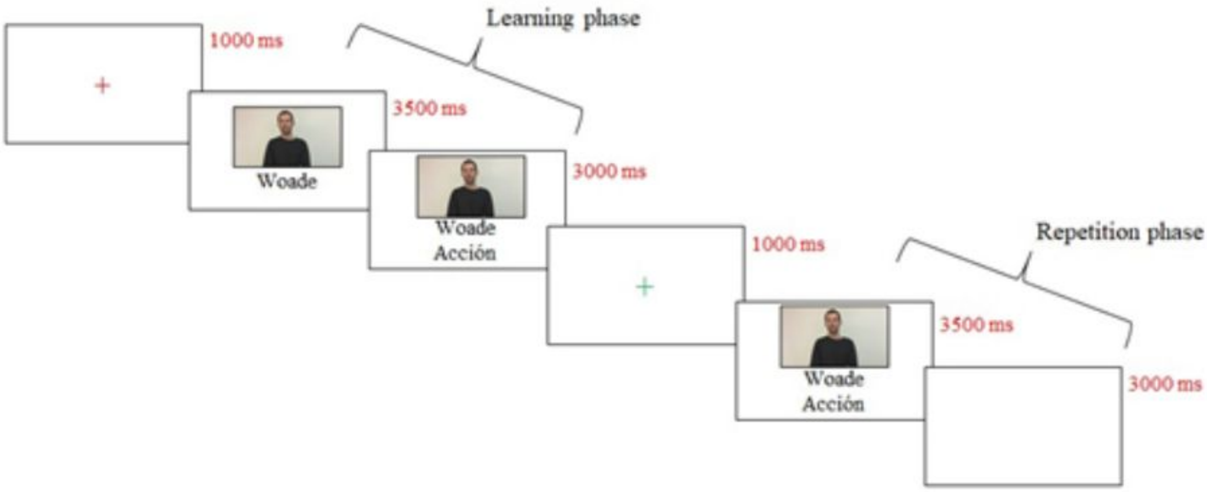

Figure 4. Procedure followed in Studies 1 and 2.

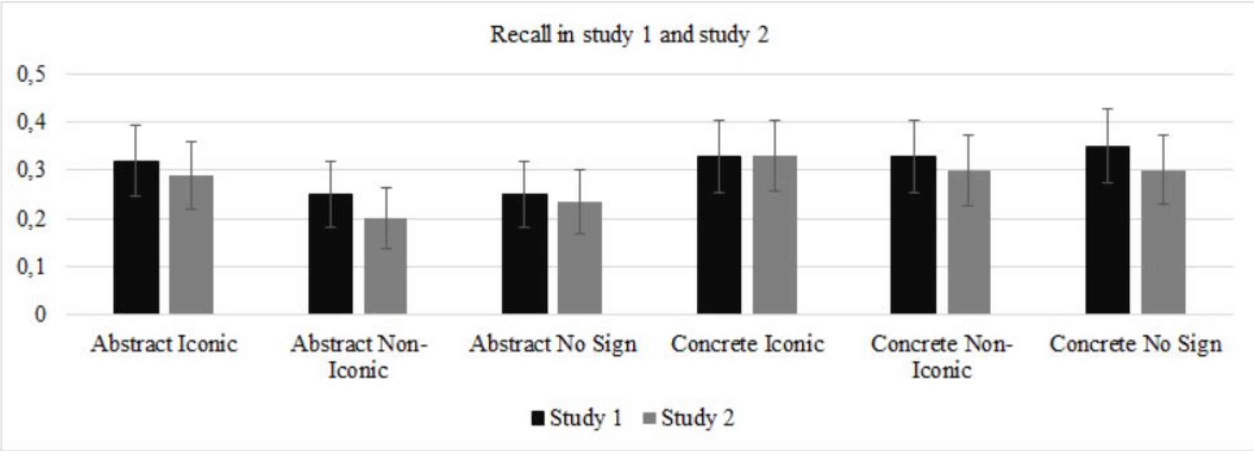

Figure 5. Recall in Studies 1 and 2.

Note. Figure 5 shows recall accuracy (from 0 to 1) for new words in the abstract iconic, abstract non-iconic, concrete iconic, and concrete non-iconic conditions in Studies 1 and 2. Error bars show standard error values.