CHAPTER VI – CONCLUSION AND SUGGESTIONS

6.1. Conclusion of Research

The results of this study support the hypothesis that multimodal flashcards are more effective than monomodal flashcards as a tool for learning the meanings of L2 concrete nouns. Multimodal learning results in the formation of multiple memory traces across independent and partially interconnected symbolic systems (i.e. verbal / image systems) and their subsystems (i.e. which correspond to different sensorimotor systems; auditory, visual, olfactory, haptic, etc.) resulting in an interconnected network of representations, whereas monomodal learning results in the formation of representations that have fewer interconnections between symbolic systems and subsystems and that are therefore more vulnerable to memory decay (Clark & Paivio, 1987, 1991; Paivio, 1986; Paivio & Csapo, 1973; Paivio & Desrochers, 1980; Paivio & Lambert, 1981).

The independence and partial interconnectedness of each system means that if one representation or connection in a network decays, the rest of the network can still function and may be able to retrieve (or perhaps even reconstruct) the required information by means of other representations and connections in the network. For example, if V2 → V1 connections are weak, then translating between L2 and L1 (i.e. translating between a V2 and V1 code) may occur by means of the Image system (e.g. V2 → I → V1). Likewise, if I → V2 connections are weak, then naming an object in L2 (i.e. translating an Image code

to a V2 code) may occur by means of the V1 system (e.g. I → V1 → V2). Such cognitive flexibility that results from dual-encoding (i.e. the availability of multiple possible routes for the retrieval of one piece of information) thus has an additive effect on recall. In contrast, monomodal learning results in the formation of representations that have fewer interconnections between symbolic systems and subsystems and that are therefore more vulnerable to memory decay. For example, if V2 → V1 connections become inviable (due to memory decay), and there are also no viable connections between V2 and the Image system (i.e. because the forming of such connections was not facilitated by the learning method, e.g. by associating an L2 word with a picture), then translating from L2 to L1 (i.e. translating between a V2 and V1 code) becomes impossible.

The benefits of multimodal (or multisensory) learning can be observed not only in response to multimodal (or multisensory) test cues (as in the study phase of the current study) but also in response to monomodal (or unisensory) test cues (as in the test phase of the current study), because the interconnectedness of each system means that stimulating one part of a network can (re-)activate the whole (a process called redintegration / spreading activation), so that – for items learned in the context of a multisensory set – retrieval operates on a richer and more informative network of interconnected representations, resulting in higher recall accuracy (c.f. Moran et al., 2013, p. 589).

It must be emphasised that multimodal learning only has a mnemonic advantage over monomodal learning under certain conditions. This study tested

one specific method of multimodal learning, and applied it to learning a specific class of L2 words i.e. concrete nouns. Therefore, it cannot be concluded from the results that any method of multimodal learning is superior to any method of monomodal learning and for any purpose, as this would vastly exceed the scope of this research. For example, providing a lot of visual information on one digital flashcard (e.g. L2 word, IPA phonemic transcription, dictionary definition, example sentence, picture of referent, video of referent, video of a mouth pronouncing the word, video of gesture, L1 word) may lead to cognitive overload and ineffective learning (encoding), which would likely result in poorer recall accuracy compared to word pairs learned from monomodal flashcards. In other words, multimodal teaching done badly is inferior to monomodal teaching done well. The author suggests that the various studies on the topic of using pictures in L2 vocabulary learning have come to a variety of conclusions because the use of pictures can help or hinder or make no difference to learning depending on exactly how the pictures are used in the learning process.

The design of the multimodal cards used in this study was informed by Mayer's principles for multimedia design in order to maximise the chance of dual-encoding while managing the risk of cognitive overload by reducing unnecessary cognitive load. These principles – which are supported by Mayer's empirical research and based on Mayer's three theory-based assumptions – may be used by educators, learners, and research designers to design effective multimodal educational materials (R. E. Mayer, 2002).

Both the strength of connections and the number of connections of one representation within a representational network contribute towards higher recall accuracy of that representation. In other words, both the quantity (number) and quality (strength) of a representation's connections within a representational network are important. On the one hand, focussing too much on forming multiple associations at once may result in the formation multiple very weak associations (connections) that quickly decay. On the other hand, focussing too much on forming one strong association (e.g. between two verbal systems) may result in a representational network that lacks the richness and depth that benefits recall and allows for greater cognitive agility and creativity. If, however, learners aim for both, then this will arguably result in the best learning outcomes. The two goals must be kept in balance, much as athletes train for both strength and flexibility. Indeed, much as different athletes who compete in different sports need differing degrees of strength and flexibility, different language learners have different learning goals and priorities that should determine which mode or modes they prioritise. For example, it would arguably be a waste of time for someone who is learning a dead language solely for the purpose of understanding and translating ancient texts to spend much time focussed on 'perfecting' their pronunciation of the language. Similarly, if someone is learning a spoken language that has no written form (which is the case for many languages, see Doner, 2024) then there is no need for them to invent a writing system for the language; they can focus on their oral comprehension and production instead.

Language is often multimodal, meaning we use multiple modalities simultaneously to communicate, such as spoken words and gestures (Doner, 2024), and language refers to objects in the real world that can be experienced through multiple senses, so it is not surprising that multimodal learning benefits L2 vocabulary learning, since the human mind is designed to process input from multiple senses simultaneously and to form associations between them.

6.2. Implications

The main implication of this study for second language teaching and learning is that creating and using multimodal digital flashcards (i.e. that follow the same design outline as those used in this study) for learning L2 concrete nouns is worthwhile because multimodal flashcards have been shown to be significantly more effective than monomodal flashcards as a tool for learning the meanings of L2 concrete nouns. Using multimodal flashcards for learning L2 vocabulary is in line with what Nation calls the dual-encoding principle – i.e. having both linguistic and nonlinguistic (e.g. pictorial) associations for a word aids word retention (Nation, 2013, p. 467).

Another implication of this study is that using digital flashcard software (such as Anki) can be an effective tool for learning L2 vocabulary, especially if cards are enhanced with semantically congruent pictures and audio like the multimodal cards used in this study. Before beginning the study participants did not know the meanings of any of the L2 words, but after 7 study sessions over the course of 7 days, the average number of successful recalled word meanings was

13.6 out of 15 (i.e. 91%) for multimodally learned words, and 12.36 out of 15 (i.e. 82%) for monomodally learned words. The effectiveness of digital flashcard software as a tool for L2 vocabulary learning has also been demonstrated by a number of other studies. To name a few examples: Nguyen found that with the support of Anki, EFL learners were able to achieve high lexical retention and retrieval (Nguyen, 2021). In Turkey, Bakla and Çekiç compared the effectiveness of using the flashcard software Memrise to using traditional vocabulary exercises for learning English vocabulary and found that there was a significant difference between the control group who used traditional vocabulary exercises and the experiment group who used Memrise, in favour of the experiment group (Bakla & Çekiç, 2017). In Makassar, Indonesia, Ervan Jaya found that using Anki was effective at improving students' English vocabulary mastery (Ervan Jaya, 2016). In Parepare, Indonesia, Mujahidah and colleagues found a statistically significant improvement in students' English vocabulary mastery resulting from the effective implementation of Anki (Mujahidah et al., 2024). Therefore, language teachers and learners can be confident that using digital flashcard software with built-in spaced repetition such as Anki can be an effective tool for learning L2 vocabulary, and enhancing flashcards with pictures and audio can make Anki even more effective as a vocabulary learning tool.

6.3. Limitations

A number of limitations of this research study can been identified. Firstly, as discussed in section 4.3.3, the results suggest that some participants may have

learnt from their incorrect answers in previous tests to improve their score in subsequent tests (Tests 2 and 3), which was possible because the test was self-marked. This could be avoided by using a blind test that is marked by the researcher (i.e. in which participants receive no feedback).

Secondly, the difference in recall of multimodally and monomodally learned words was, on average, smaller for Group A participants (who learned List 1 vocabulary multimodally and List 2 vocabulary monomodally) than for Group B participant (who learned List 2 vocabulary multimodally and List 1 vocabulary monomodally), which suggests that List 1 vocabulary was more difficult to learn than List 2 vocabulary. This word difficulty bias could be avoided by empirically testing the difficulty of each word pair before commencement of the study with a representative sample of participants from the same population as the research study participants. The difficulty of the vocabulary could then be balanced between the two sub-lists based on this empirical data, rather than based on a linguistic analysis of word complexity alone.

Thirdly, it seems that length of Test Phase not sufficient to observe memory decay between post-tests. Recall accuracy did not significantly decrease in tests 2 and 3, which suggests that the length of time that elapsed between post-tests was not sufficient for memory decay to occur between post-tests. This could be rectified by increasing the length of time between post-tests. If tests 2 and 3 had been carried out following a much longer interval, this would have increased the likelihood of participants forgetting previously learnt words (memory decay),

and we may have been able to observe if the meanings of L2 words learned in a multimodal way were retained (remembered) for longer than the meanings of L2 words learned in a monomodal way.

Fourthly, the validity of this study depends upon research participants' ability and willingness to follow the researcher's instructions. Participants were asked to tap *Ulang* 'Again' if they failed to recall an answer correctly and *Baik* 'Good' if they recalled an answer correctly. Because participants' answers were self-marked, it was possible for participants to tap 'Good' even if they failed to recall the answer correctly, however it is difficult to imagine a motive for doing so since participants were not aware of the research objectives, they knew that their performance data would be anonymised, and they had agreed to follow the researcher's instructions.

Lastly, it should be noted that some word pairs were reviewed more times than others. During the study phase, word pairs that were marked 'Again' remained in the learning queue and came up for review again within the same study session, meaning that some word pairs (i.e. ones that participants found more difficult to recall) were reviewed a greater number of times than others. This design was intentional so that it would reflect real-life learning with Anki, however this did mean that the first post-test (Test 1) was not a completely fair test since – on average – monomodally learned word had been reviewed a greater number times prior to the test. It is remarkable then that despite the fact that – on average – monomodal cards were reviewed more often than multimodal cards

during the study phase (which will have aided subsequent recall), recall accuracy for multimodally learned word pairs in Test 1 was still significantly better than recall accuracy for monomodally learned word pairs. It is likely that, had multimodally and monomodally learned word pairs been reviewed the same number of times, the effect size (which was medium) would have been even larger.

6.4. Suggestions for future research

This study tested one specific method of multimodal learning, and applied it to learning a specific class of L2 words i.e. concrete nouns. Future studies could test different monomodal and multimodal card designs, e.g. multimodal cards that incorporate gesture (video), and monomodal cards that are include only auditory words. Future research could even compare the effectiveness of 'No L1' flashcards (e.g. L2 + picture) and 'Three-system' flashcards (e.g. L2 + picture + L1) for L2 vocabulary learning. For a summary of all the possible stimuli that could be utilised in multimodal and monomodal L2 vocabulary learning, see Table 21 below. Some modes listed in Table 21 (e.g. the smell, taste and feel of objects) cannot be incorporated into digital flashcards due to current limitations in technology, but may be incorporated into classroom teaching by utilising real physical objects that can be seen, touched, smelt, and even tasted.

Table 21: *A summary of the possible stimuli that could be utilised in multimodal and monomodal L2 vocabulary learning.*

System	Subsystem	Multimodal (choose any combination of modes)	Monomodal (choose one mode)
Verbal 1, 2 etc.	Visual	Visual words	Visual words/
	Auditory	Auditory words	Auditory words/
	Motor	Signs	Signs
	Haptic (Touch)	Writing / typing patterns; braille	
Image (nonverbal)	Visual	Pictures	
	Auditory	Environmental sounds	
	Motor	Gestures	
	Haptic (Touch)	"Feel" of objects	
	Taste	Taste of objects	
	Smell	Smell of objects	

The current study tested receptive retrieval only (L2 → L1), but future studies could test productive retrieval (L1 → L2) instead of or in addition to receptive retrieval. This study only used concrete nouns, but future studies could use other parts of speech (e.g. verbs, adjectives, adverbs, abstract nouns, conjunctions) and could investigate whether using pictures benefit the learning of some parts of speech more than others.

This study tested word retention 1, 2 and 4 days since most recent exposure (see Table 12), but future research could test word retention following a longer interval (e.g. in a post-test weeks or months after completion of the study

phase) – increasing the likelihood of participants forgetting previously learnt words (memory decay) – to test if the meanings of L2 words learned in a multimodal way are retained (remembered) for longer than the meanings of L2 words learned in a monomodal way.

This study measured the number of user-initiated reviews (‘Again’ count) during the study phase and the number of correct recalls (‘Good’ count) during the test phase, but future research could measure a different dependent variable to investigate other potential benefits of multimodal learning over monomodal learning, e.g. benefits for pronunciation accuracy, spelling accuracy, auditory word recognition, and object naming.

Finally, all the participants in this study were Indonesian adults, but future research could attempt to test the hypothesis of this study with child participants.