

# Evaluating types and combinations of multimodal presentations in the retention and transfer of concrete vocabulary in EFL learning<sup>1</sup>

*Evaluación de tipos y combinaciones de presentaciones multimodales en la retención y la transferencia de vocabulario concreto en el aprendizaje de EFL*

Miguel Farías

UNIVERSIDAD DE SANTIAGO DE CHILE  
CHILE  
miguel.farias@usach.cl

Roxana Orrego

UNIVERSIDAD DE SANTIAGO DE CHILE  
CHILE  
roxana.orrego@usach.cl

Katica Obilinovic

UNIVERSIDAD DE SANTIAGO DE CHILE  
CHILE  
katica.obilinovic@usach.cl

Tammy Gregersen

UNIVERSITY OF NORTHERN IOWA  
USA  
tammy.gregersen@uni.edu

**Recibido:** 1-V-2012 / **Aceptado:** 3-VI-2013

## Abstract

Although multimodality has been addressed in the educational psychology arena, this study evaluates the effects of presentation modality in the retention and transfer of concrete vocabulary in students of English as a second language. The participants were 104 second year university students belonging to three groups. Group 1 (n=32) was exposed to vocabulary through on screen text and narration (Group TN); group 2 (n=42) was exposed to on screen text, narration and video (Group TNV); and group 3 (n=30) was presented vocabulary via on screen text, narration and still image (Group TNI). All three groups were given a retention test (RT), a transfer test (TT), and a questionnaire to evaluate the type of presentation (TPQ). Results indicate that in the RT there are statistically significant differences among groups (ANOVA), being the TNI group the one that retains more lexical items. In the TT, using the t-student, results show that there are significant differences between groups TNI and TN, being group TNI the one with the highest scores. Results from the TPQ suggest that still images helped more than text and video in vocabulary learning, that actions are better represented through videos than through still images, and that more attention is paid to narration in group TN than in groups TNV and TNI.

**Key Words:** Multimodality, redundancy principle, retention, transference, SLA.

## Resumen

Este estudio es uno de los primeros que evalúa los efectos de la modalidad de presentación, en la retención y la transferencia de vocabulario concreto, en los estudiantes de inglés como segunda lengua. Los participantes fueron 104 estudiantes universitarios de segundo año pertenecientes a tres grupos. El grupo 1 ( $n = 32$ ) fue expuesto a vocabulario en una presentación con texto en pantalla y narración (Grupo TN); el grupo 2 ( $n = 42$ ) fue expuesto a texto en pantalla, narración y vídeo (Grupo TNV); y al grupo 3 ( $n = 30$ ) a texto en pantalla, narración e imagen fija (Grupo TNI). Los tres grupos rindieron una prueba de retención (RT), una de transferencia (TT), y un cuestionario para evaluar el tipo de presentación (TPQ). Los resultados indican que en la RT existen diferencias estadísticamente significativas entre los 3 grupos (ANOVA), siendo el grupo TNI el que más retiene elementos léxicos. En la TT, utilizando la *t-student*, los resultados muestran diferencias significativas entre los grupos TNI y TN, siendo TNI el grupo que obtiene los mejores resultados. Los resultados del TPQ sugieren que las imágenes fijas ayudan a retener y transferir información de mejor forma. Asimismo, las acciones estarían mejor representadas a través de videos que de imágenes fijas, y que se presta mayor atención a la narración en el grupo TN que en los grupos TNV y TNI.

**Palabras Clave:** Multimodalidad, principio de redundancia, retención, transferencia, SLA.

## INTRODUCTION

The introduction of the concept of multimodality into applied linguistics has provided valuable constructs that help account with greater certainty for the learning process in multimodal contexts. Despite the fact that formal language learning has traditionally used multimodal aids, the theoretical reflections coming from both research in multimodal learning within the field of learning psychology (Mayer, 2001; Schnotz, 2002, 2005) and the sociosemiotic models of multimodal text design and production (Kress & van Leuween, 1996, 2001; Kress, 1997, 2000) have renovated research interest in understanding how language presented in multimodal formats is comprehended and produced. Even though there has been substantial research on multimodal learning (Mayer & Sims, 1994; Mayer & Moreno, 1998; Mayer, Heiser & Lonn, 2001), the content areas or disciplines involved in such research do not share the intrinsic characteristics of second language acquisition (SLA) thus making empirical evidence in the SLA field imperative, more so given the ever-increasing exposure to multimodal texts by SL learners.

Such renovated interest is triggered by the urgency to grapple with and, ideally, to forecast an understanding of how language is processed in the context of rapidly changing information and communications technologies (ICT's). For some authors (Unsworth, 2006; McDonald, 2009; Serafini, 2011) this complex scenario of emerging

language presentation modalities calls for the creation of a metalanguage that can guide our understanding of the possible relations and interactions between text and image. Still for others, such metalanguage is a key component in developing the multiliteracies that our contemporary societies require (Cope & Kalantzis, 2000; Kress, 2003). One of the skills often mentioned is visual literacy (Fariás & Orrego, 2008).

Teacher educators in particular need to explore the potential that multimodality can offer second or foreign language learning (Fariás, Obilinovic & Orrego, 2007, 2011). Plass and Jones (2005) have contributed to the dialog of second language learning and multimodality by postulating a model that integrates these two fields. In their analysis they adopt Mayer's model of multimedia learning and they take elements from both Chapelle's and Ellis's SLA models. The essential question that motivates their proposal and also is an impetus in ours is: "In what way can multimedia support second-language acquisition by providing comprehensible input, facilitating meaningful interaction, and eliciting comprehensible output?" (Plass & Jones, 2005: 471).

Along the same lines, Mayer (2001) includes a set of multimodal learning principles that can be used in answering this question. In his model learning is measured in terms of two processes: 'retention' and 'transfer'. Information retention is the preliminary and obligatory step to achieve transfer; that is, the application of such information in new contexts. However, Mayer (2001) places more emphasis on transfer given the fact that it provides evidence of meaningful learning to the extent that transfer can account for the ability to apply the contents that have been stored in memory. These are the two types of learning to be measured in this study.

## **1. Literature review**

Until the arrival of the Lexical Approach (Lewis, 1993), the teaching of vocabulary had played a minor role in the different methods used to teach a second language (Zimmerman, 1997). Vocabulary acquisition became the focus of attention in theories of language learning (McCarthy, 1990; Lewis, 1993; Coady & Huckin, 1997; Read, 2000; Schmitt, 2000; Nation, 2001) and in terms of the vocabulary learning strategies students needed to use to attain an effective pool of language for communicative purposes (Gairns & Redman, 1986; Carter & McCarthy, 1988; Schmitt, 1997; Sökmen, 1997). Vocabulary acquisition is here reviewed in studies dealing with second language multimodal learning (Plass, Chun, Mayer & Leutner, 1998; Dubois & Vial, 2000; Fariás, Obilinovic & Orrego, 2009; Syrodenko, 2010).

In the wide ever-growing field of multimodal learning one of the basic questions concerns how the learning process can be made more beneficial for learners when the teaching materials include not only text but also images and narration (also called teacher's voice or input in SLA studies). Of the multimodal learning principles

mentioned, the one that poses the greatest challenge for SLA is the redundancy principle (Mayer, 2001). The term redundancy effect was introduced by Kalyuga, Chandler and Weller (1998: 2) to address situations in multimedia presentations in which “eliminating redundant material results in better performance than when the redundant material is included”. Mayer (2001: 153), however, uses the term as we understand it in this paper, i.e. “any multimedia situation in which learning from animation (or illustrations) and narration is superior to learning from the same materials along with the printed text that matches the narration”. The redundancy principle states that people learn better from animation and narration (teacher’s voice in this study), than from animation (still image and video in this study), narration (teacher’s voice), and on-screen text (written text in the PowerPoint presentations). Notwithstanding, how redundancy is understood depends on the conceptions of learning the researcher holds. A traditional view of learning, and certainly one shared by most language teachers, would hold that presenting words in two sensory modalities is better than presenting words in only one way. In this respect, we have argued (Fariás, Obilinovic & Orrego, 2009, 2013) that for second language beginner learners presentations including spoken and printed words may prove beneficial as text acts as subtitle to a movie until enough proficiency is attained to rely on just one mode. Additionally, we need to mention the learning-preferences hypothesis, under which view having two modalities available would cater for different learning styles, i.e., the visual learner would focus on the image and the auditory learner on narration. Even though research in other fields have not confirmed redundancy as an efficacy enhancing factor for learning, it may work differently in SLA by helping learners retain and transfer the new linguistic code. The distinction between these two areas of learning has been explored by Schnotz and Baadte (2008: 22) as they claim that “in language learning, things are different because the primary goal of learning is not to learn about a specific domain, but to master a new language”. Thus, Schnotz and Baadte (2008) make the difference between language learning and domain learning where the major distinction is that language learning precedes domain learning, understood as the kind of learning that is institutionalized in the school system: biology, history, geography, etc. On the other hand, in second language learning (particularly in adolescents and adults) the learner already possesses knowledge of the domain (unless it is highly specialized), in which case the multimodal presentation of the target language can be either in the form of a meaningful context to activate prior knowledge and trigger an understanding of the new code, or as schematized formal rules that engage conscious learning of aspects of the target language system (syntactic, prosodic or semantic rules) (following the distinction between conscious learning and unconscious acquisition postulated by Krashen & Terrel, 1983). This study compares three types of multimodal second language presentations aimed at providing a meaningful language environment for vocabulary acquisition through the use of text, voice, still and moving images.

Plass et al. (1998) carried out previous studies on multimodal learning investigating English speaking learners of German in a multimedia learning environment. These learners read a text in German that was presented to them through software that offered them annotated translations of key lexicon, a video clip illustrating the lexical item, or both. The results showed that learners did better when they used verbal and visual annotations and that they comprehended the story better when they used their preferred mode of annotation. These two types of learning reported by these authors are the ones under investigation in this study: retention and transfer.

Similarly, Plass (1998) has investigated the interfaces utilized in the interaction with multimedia in foreign language learning software. The author presents four approaches used for describing the design of user interface (craft, enhanced software engineering, technologist and cognitive) and explains that he has adopted the cognitive model as the most appropriate for the design and evaluation of user interface computer programs for second language acquisition due to the fact that such an approach can incorporate in its design both the user as well as the learning task. In the present study, the cognitive approach was used in the three types of presentations as participants used their listening and reading comprehension skills to understand language presented through the teacher's narration and the written text.

Dubois and Vial (2000) investigated the interaction between verbal and visual modes of presenting the foreign language for Russian vocabulary learning by French speaking learners. They predicted that their students would show better recall when textual information was presented with visual and auditory information as with semantic and phonetic links between the elements. Their theoretical rationale for this claim was that "when textual, visual, and auditory materials are integrated in this way, the learner may be forced to engage in additional processing that leads to better memorisation" (Dubois & Vial, 2000: 159). Among the results obtained, Dubois and Vial (2000) explain that auditory information presented together with visual elements fostered more learning than textual information presented with the same image, thus confirming the findings of other authors like Mayer and Moreno (1998). Dubois and Vial (2000: 163) explain the consistency of the results saying that "a presentation where both elements presented (image and text) are only visual results in less learning than if both are visual and auditory (less cognitive load)". In Mayer's terminology, this coincides perfectly well with his modality principle which states that "students learn better from animation and narration than from animation and on-screen text" (Mayer, 2001: 184). This cognitive load resulting from some types of presentations is associated to the redundancy effect when an extraneous load is introduced and information has to be processed in working memory by the same channel (visual: image and text) rather than by two channels (visual and auditory: image and narration). In Farías et al. (2009) we investigated the effects of two types of presentation in the retention and transfer of idiomatic expressions in an EFL context, one including narration, text and image and another only narration and text. Although there were no differences between groups, the discussion centered on the nature of the language

presented since idiomatic expressions carry figurative or metaphoric meanings that can be hardly iconically represented in images. Consequently, for this study we have chosen concrete vocabulary items.

Furthermore, and particularly important to the present study, Syrodenko (2010) investigated the effects of modalities of input on vocabulary acquisition of Russian as a foreign language. Acquisition was measured via written and aural vocabulary tests. Attention to input and use of learning strategies were measured by a questionnaire. Results indicate that groups exposed to video with audio and captions and to video with captions scored higher on written than aural recognition of vocabulary and that the group exposed to video and audio scored higher on aural than written word recognition. In the questionnaire, learners mentioned paying most attention to captions, then to videos and then to audio and that they learned most words by associating them with visual images.

In this research tradition, this paper reports on a study aimed at evaluating the effects of three types of multimodal presentations on the retention and transfer of concrete vocabulary of English as a second language by Spanish-speaking Chilean university learners. Three presentation formats were used in this quasi experimental study: a) TN: on-screen text and narration, b) TNI: on-screen text, narration and still images; and c) TNV: on-screen text, narration and videos.

## **2. The study**

Participants in this study included three groups of university students, two from an English teacher education program and one from a translation program. Two of the groups are from a state university and one from a private university. Learners in the three groups were in their second year of studies and their level of English corresponds to pre-intermediate, according to the ALTE scale. The number of hours dedicated to the English language instruction in their first year of studies is the same in both institutions and in both programs.

### **2.1. Research questions**

1. Are there significant differences in retention among groups of learners exposed to three different formats of vocabulary?
2. Are there significant differences in transfer among groups of learners exposed to three different formats of vocabulary?
3. How do learners perceive the contribution that text, narration, still images and videos make to their vocabulary acquisition?

### **2.2. Procedures**

The study involved four stages: a diagnostic pre-test, exposure to three types of presentations (TN, TNI y TNV), post tests (retention and transfer) and questionnaire

application. Groups were already formed so the decision was made to assign the two groups in teacher education programs (from two different universities) to the TN and the TNV groups, respectively, and the translation studies group to the TNI presentation.

### **2.3. Instruments**

- a) **Pre-test:** A list of 30 low frequency verbs in English was created, which included lexical items of a concrete nature, able to be iconically depicted. This 30 item test contained multiple choice answers and was used to diagnose the vocabulary pool of the participants. (See Appendix 1)
- b) **Retention test:** The most unknown vocabulary items from the pretest were used both for the presentations and to create a multiple choice retention test. Following Miller's hypothesis of a limited capacity of around seven objects to be held in short term memory (Miller, 1956) thirteen items were selected for this test: the ten that were included in the presentations plus three that acted as distractors. This makes a possible maximum score of 10 points. Results were calculated without the three distractors. (See Appendix 2)
- c) **Transference test:** The ten items in the presentations were included in this test but in communicative contexts where the respondents had to complete a sentence with the appropriate verb. Care was taken so that the level of the language used in creating this communicative context matched the proficiency level of the participants. Each correct answer obtained 2 points. Answers that were misspelled but that kept the meaning of the word obtained 1 point. The maximum possible score was 20 points. (See Appendix 3)
- d) **Type of presentation questionnaire:** Using Syrodenko's (2010) procedures, a questionnaire was adapted for each of the three groups to measure how learners evaluated the different presentation modalities for their contributions to their vocabulary acquisition. These three questionnaires contained some Likert-scale answers ('Yes, totally agree'; 'Yes, to some extent'; 'I don't think they helped me at all') as well as open-ended questions. The Likert-scale responses were tabulated using percentages while the open-ended questions used semantic groupings.

### **2.4. Presentations**

The ten verbs that were, according to the diagnostic test, the least known by the three groups of learners were included in the three types of presentations. Using PowerPoint, these presentations included the following elements (here as an example illustrated with one of the ten verbs: to sow):

1. TN: on-screen text containing the verb and a brief definition of its meaning (*To sow*: 'To sow is to spread seeds in the ground, which will grow some day and become plants') and narration (a voice that reads the same text).

2. TNI: on-screen text containing the verb and brief description of the action that is portrayed in the image (“To sow. This man is sowing. He is spreading seeds in the ground, which will grow some day and become plants”), a voice reading the on-screen text and a still image illustrating the action of the verb.
3. TNV: on-screen text of the verb (To sow), narration as voice describing the action illustrated in the video and the moving image as video: ‘This woman is sowing. She is spreading seeds in the soil which will grow some day and become plants’.

In the three presentations no translation was used, ie, there were no subtitles with L1 text and L2 sound, nor reversed subtitles with L1 sound and L2 text but captions: text and sound in the target language.

These presentations included not only a comparison between on-screen text plus narration and a presentation based on the redundancy principle (Mayer) with narration and the double use of the visual channel for processing both the text and the image, but also a comparison between two types of images: still and moving.

**Table 1.** Three presentation modalities and number of participants.

Group TN n=32	Group TNI (redundant) n= 30	Grupo TNV (redundant) n=42
On-screen text	On-screen text	On-screen text
Narration (or teacher’s input)	Narration (or teacher’s input)	Narration (or teacher’s input)
	Still images	Moving images (video)

Table 1 above shows the characteristics of the three presentations.

### 3. Results

#### a) Retention

Table 2 below illustrates the correct answers in the three groups, 10 being the maximum. The average in the three groups was 8.78 points. Group TNI scores above the general average obtaining 9.13 points and a minimum of 6 correct answers. On the other hand, the same group reaches the maximum of 10 correct answers. The standard deviation of 0.97 is the lowest of the groups.

Group TNV scores an average of 8.88 correct answers, with a median of 9 points, which corresponds to 50% of the cases around this value. This group shows a minimum of 3 correct answers but reaches the maximum of 10 points. The mode is also 10 points and the standard deviation is 1.38.

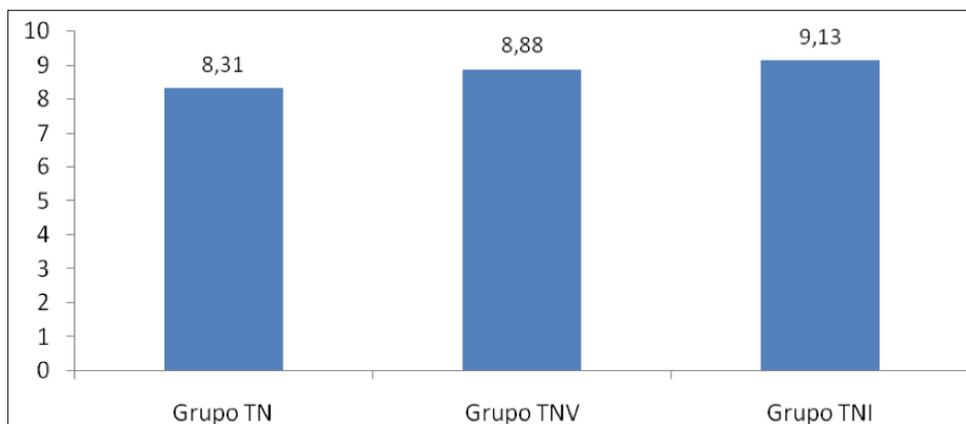
Finally, group TN scores the lowest amount with a total of 8.31 points. The minimum score is 4 and the maximum is 10 with a standard deviation of 1.51 and

the mode 8 points. We conclude that this group shows the lowest number of correct answers.

**Table 2.** Correct answers in retention test in the 3 groups.

	Group TN	Group TNV	Group TNI
N°	32	42	30
Mean	8.31	8.88	9.13
Median	8	9	9
Mode	8	10	9
Minimum	4	3	6
Maximum	10	10	10
Standard deviation	1.51	1.38	0.97

Graph 1 below shows the results from the retention test by highlighting the average in the correct answers and showing that Group TNI is the one with the best results of the three groups.



**Graph 1.** Mean of correct answers in the retention test in the 3 groups.

The following hypotheses were formulated in applying the ANOVA test to evaluate significant differences among groups:

- **Ho:** There are no statistically significant differences in the results from the retention test applied to the three groups. The means are the same for the three groups.
- **Ha:** There are statistically significant differences in the results from the retention test applied to the three groups. The means are different for the three groups.

Results from the ANOVA using the SPSS demonstrate that with a 95% confidence level and an error margin of +/- 5% there are statistically significant differences<sup>2</sup> between groups TN and TNI in the number of correct answers: group TNI retains more than group TN. No significant differences were found between group TNV and the other groups.

### b) Transference

Table 3 shows the results of the correct answers in the three groups. The average number of correct answers in the three groups is 10.78 out of a total of 20.

Group TNI goes over the average by reaching an average of 12 correct answers with a minimum of 4. This group has a maximum of 19 correct answers with a standard deviation of 4.61.

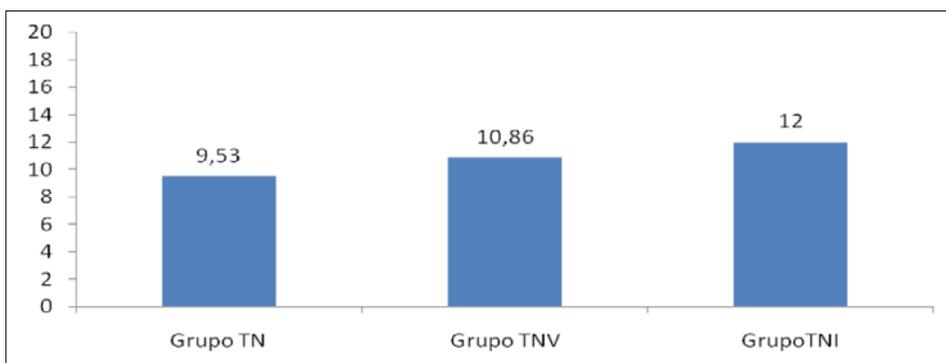
Group TNV's average is 10.86 correct answers, with a median of 12.

Group TN earns the lowest score with an average of 9.53 correct answers and a standard deviation of 3.84.

**Table 3.** Correct answers in transference test in the 3 groups.

	Group TN	Group TNV	Group TNI
N°	32	43	30
Mean	9.53	10.86	12
Median	9	12	14
Mode	9	12	14
Minimum	3	0	4
Maximum	18	20	19
Standard Deviation	.84	5.89	4.61

Graph 2 shows the means in the three groups, where Group TNI presents the best general evaluation and the means fluctuate between 9.53 and 12 correct answers.



**Graph 2.** Mean of correct answers in the transference test in the 3 groups.

To compare the means using the T- student, the following hypotheses were raised:

- **Ho:** There are no statistically significant differences in the results from the transference test applied to the three groups. The means are the same for the three groups.
- **Ha:** There are statistically significant differences in the results from the transference test applied to the three groups. The means are different for the three groups.

Results from the t-student, using the SPSS, indicate that with a 95% confidence level and a margin of error of +/- 5% there are statistically significant differences<sup>3</sup> between groups TN and TNI in the scores for correct answers: Group TNI transfers more than group TN. There are no statistically significant differences between groups TN and TNV, nor between groups TNI and TNV.

### c) Type of Presentation Questionnaire (TPQ)

The results presented in continuation illustrate the qualitative feedback provided by the students to the following questions:

1. Do you believe that (the text-TN, the video-TNV, the images-TNI) contributed to your understanding of the words being presented?

**Table 4.** Results from the question on the contribution of text, video and still images in understanding vocabulary.

Answer	Group TN	Group TNV	Group TNI
Yes, totally agree	28.6%	55.8%	66.7%
Yes, to some extent	71.14%	41.9%	33.3%
I don't think they helped at all	0%	2.3%	0%
Total	100%	100%	100%

Table 4 shows the results to the question in the TPQ that asked learners to evaluate the contribution of text, video and still image to their understanding of vocabulary. Group TNI attaches more value to the images as a contributor to their understanding of vocabulary (66.7%); in turn, group TNV assigns 55.8% to the video as a contributing factor and Group TN a 28.6% to the text. Interestingly, participants in Group TN also mentioned that the text helped them “to some extent” (71.14%), which confirms the perception that text, along another mode –narration, was a contributing factor in understanding the vocabulary presented.

The second question presented for qualitative evaluation was the following:

2. While I was reading the text/watching the video/looking at the image, I was paying attention to the narrator’s voice.

**Table 5.** Attention to narration.

Answer	Group TN	Group TNV	Group TNI
All the time	64.3%	48.8%	46.7%
At times	35.7%	51.2%	53.3%
Total	100%	100%	100%

Results from other questions in the TPQ indicate, as shown in Table 5, that the attention paid to the narrator’s voice in Group TN is the highest with a 64.3%, followed by group TNV with 48.8% and by Group TNI with 46.7%. These results on how narration contributes to vocabulary learning in Table 5 show an interesting opposite relation to those in Table 4: more attention is paid to narration when it is accompanied only by text than when it comes with text and images. By not having images available there is no redundancy effect as narration represents for Group TN the obligatory and most natural complement to the textual presentation of language, whereas for groups TNI and TNV text and image compete (and overload) to be processed by the visual-pictorial channel.

On the other hand, when the actions were illustrated both through still and moving (video) images slightly greater attention was paid to narration by group TNV than by group TNI. Higher ‘At times’ than ‘All the time’ answers in groups TNV and TNI may indicate that in selecting and organizing information that has been coded visually twice (printed text and image), narration played a secondary role and was needed at times for confirmation.

Participants in Group TN mentioned, in the open-ended questions of the TPQ, that they would have appreciated having images included in the presentation and recommended including actions as complement to the texts.

Regarding the question on the problems they faced during the presentations, respondents mentioned that the words were too far from the image and the text, that most of the items were monosyllabic and some were very similar, for example, row and sow; that the images were distracting and that some images were not clear enough.

## 4. Discussion

Previous investigations in educational psychology that targeted the redundancy principle do not necessarily concur with our findings (Mayer & Sims, 1994; Mayer & Moreno, 1998; Mayer et al., 2001), particularly in reference to moving images. One of the problems in designing the presentations was the need to have clear moving images that illustrated the verbs without confusing distracters. Despite the fact that the verbs selected represented concrete actions that were potentially easy to be portrayed through moving and still images, these images invited several interpretations on the part of the learners as they did not know the exact meaning being conveyed. That is the case, for example, of the verbs 'to stir' and 'to hem' that were also interpreted as 'to cook' and 'to sew', causing respondents to answer incorrectly. This intrinsic difficulty in finding explicit iconicity between image, text and narration may be resolved by resorting to translation in the mother language to make sure that the learner comprehends the intended meaning of words that may have several possible interpretations.

An additional comment is needed here regarding the manner in which participants were exposed to text, still and moving images. The presentations containing still images included text, defined not only as the infinitive form of each verb shown on screen on the upper part of the PowerPoint slide but also the literal script of the oral narration that participants were listening to and that was included at the bottom of each still image. The presentations with video, on the other hand, only included the text of the infinitive form of the verb in the upper part of the slide while the video was shown on the lower part and the voice of the narrator described the action being performed. No script from this narration was included in the slide, which may explain why participants in this group did not score as high as the group exposed to the script of the narration as on-screen text (Group TNI).

A variety of input modalities provide unique benefits for different skills and styles preferences. Though short in scope for the size of the sample, this study found that the still image superseded the video in vocabulary retention and transfer in the target language. Perceptual and cognitive arguments may be raised to explain such difference. On one hand, drawing on Paivio's Dual Code Theory (Paivio, 1971), the fixed permanence of the still image as visual code, may match more effectively the verbal code, as the moving image may be more volatile in serving as pairing anchor to the verbal input. In this respect, the learner exerts more control over a still image than

a moving image, both in retaining the meaning by matching the verbal and visual code and applying it later in meaningful contexts of occurrence (transfer).

On the other hand, the weight of the pedagogical tradition may be rooted in the learners' sociocognitive system to privilege still images in constructing their personal 'pictionaries' of the target language. Even though the video may quite effectively portray the action of the verb's meaning, only one photogram of the moving image is used to store such lexical item in the pictorial model of the working memory (see Mayer, 2001). Here is where the concept of design postulated by Kress (2000, 2003, 2008) gains importance in terms of the sociocultural, cognitive and pedagogical implications involved in the construction and interpretation of multimodal texts.

Moreover, since retention and transfer were measured with written tests, such task required retrieval from the static visual text rather than from the moving image or the narration. Had the testing task been aural, it is possible that narration could have played a more dominant role as a more fluid modality where the sound can be directly associated to the word.

## **CONCLUSIONS**

The consistently higher results by Group TNI over Group TN confirm the basic premise of multimodal learning on the advantage of presentations including text and image by means of which learners actively integrate textual and pictorial information into a coherent mental model. Higher results of both TNI and TNV Groups over Group TN may indicate that TNI and TNV presentations offered more options for different learning styles preferences.

What is interesting in the results is that group TNI has scored higher than group TNV given the fact that one should expect the moving image to portray more effectively the action verbs that were presented.

Redundancy does not seem to be a problem for second language learning, at least for vocabulary learning in learners with beginner and pre-intermediate proficiency levels. Consistently, it was the group exposed to text, narration and still image that scored the highest results in retention and transfer of concrete vocabulary items. Redundancy, as the duplication of information via the same visual channel, text and image, helped learners to retain and transfer the lexical items being presented. This seems to be the case primarily in beginning and pre-advanced second language learners that require the text, the voice and the image to construct their lexicons in the second language as they have not yet attained automaticity in matching sound and text (Farías et al., 2009, 2013).

Future research may consider including a repetition of the presentations to ascertain whether a second or third exposure may serve as reinforcement in

vocabulary acquisition and/or some kind of delayed testing to measure long-term retention. Furnishing translations and/or zooming in may also help in disambiguating meanings in text-image relations that are too general.

## REFERENCES

- Carter, R. & McCarthy, M. (1988). *Vocabulary and language teaching*. New York: Longman Inc.
- Coady, J. & Huckin, T. (1997). *Second language vocabulary acquisition. A rationale for pedagogy*. Cambridge: Cambridge University Press.
- Cope, B. & Kalantzis, M. (Eds). (2000). *Multiliteracies. Literacy learning and the design of social futures*. London: Routledge.
- Dubois, M. & Vial, I. (2000). Multimedia design: The effects of relating multimodal information. *Journal of Computer Assisted Learning*, 16, 157-165.
- Fariás, M., Obilinovic, K. & Orrego, R. (2007). Implications of multimodal learning models for foreign language teaching and learning. *Colombian Applied Linguistics Journal*, 9, 174-199.
- Fariás, M. & Orrego, R. (2008). *Developing critical digital literacy in Chilean language education*. Proceedings from ACE: World Conference on Educational Multimedia, Hypermedia and Telecommunication, Chesapeake, VA.
- Fariás, M., Obilinovic, K. & Orrego, R. (2009). *Estudio exploratorio de aplicación del principio de redundancia en el aprendizaje de expresiones idiomáticas en lengua extranjera*. Actas del Segundo Congreso Nacional de la Cátedra UNESCO, Universidad de Los Lagos, Osorno.
- Fariás, M., Obilinovic, K. & Orrego, R. (2011). Engaging multimodal learning and second/foreign language education in dialogue. *Trabalhos de Lingüística Aplicada*, 50(1), 133-151.
- Fariás, M., Obilinovic, K. & Orrego, R. (2013). El principio y efecto de redundancia en la retención y transferencia de expresiones idiomáticas en inglés como lengua extranjera. *IKALÁ, Revista de Lenguaje y Cultura*, 18(1), 9-17.
- Gairns, R. & Redman, S. (1986). *Working with words: A guide to teaching and learning vocabulary*. Cambridge: Cambridge University Press.
- Kalyuga, S., Chandler, P. & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors*, 40(1), 1-17.
- Krashen, S. & Terrel, T. (1983). *The natural approach: Language acquisition in the classroom*. Oxford: Pergamon Press.
- Kress, G. (1997). Visual and verbal modes of representation in electronically mediated communication: The potentials of new forms of text. In I. Snyder (Ed.), *Page to screen* (pp. 51-53). London: Routledge.

- Kress, G. (2000). Design and transformation: New theories of meaning. In B. Cope & M. Kalantzis (Eds.), *Multiliteracies: Literacy learning and the design of social futures* (pp. 153-161). London: Routledge.
- Kress, G. (2003). *Literacy in the new media age*. New York: Routledge.
- Kress, G. (2008). Miguel Farías entrevista a Gunther Kress. In M. Farías & K. Obilinovic (Eds.), *Aprendizaje multimodal/Multimodal learning* (pp. 15-21). Santiago de Chile: PUBLIFAHU USACH.
- Kress, G. & van Leeuwen, T. (1996). *Reading images: The grammar of visual design*. London: Routledge.
- Kress, G. & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. London: Edward Arnold.
- Lewis, M. (1993). *The lexical approach: The state of ELT and a way forward*. Hove, UK: Language Teaching Publications.
- McCarthy, M. (1990). *Vocabulary*. Oxford: Oxford University Press.
- McDonald, P. (2009). *To what extent can defining graphic/written text relations support the teaching of reading comprehension in multimodal texts?* Master of Arts Thesis University of Birmingham, Edgbaston, Birmingham, Scotland.
- Mayer, R. (2001). *Multimedia learning*. Cambridge: Cambridge University Press.
- Mayer, R. & Sims, K. V. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, 86(3), 389-401.
- Mayer, R. & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual-processing systems in working memory. *Journal of Educational Psychology*, 90(2), 312-320.
- Mayer, R., Heiser, J. & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, 93, 187-198.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, and Winston.

- Plass, J. (1998). Design and evaluation of the user interface of foreign language multimedia software: A cognitive approach. *Language Learning & Technology*, 2(1), 40-53.
- Plass, J., Chun, D. M., Mayer, R. & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second language multimedia learning environment. *Journal of Educational Psychology*, 90, 25-36.
- Plass, J. & Jones, L. (2005). Multimedia learning in second language acquisition. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 467-488). New York: Cambridge University Press.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Serafini, F. (2011). Expanding perspectives for comprehending visual images in multimodal texts. *Journal of Adolescent & Adult Literacy*, 54(5), 342-350.
- Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 109-121). Cambridge: Cambridge University Press.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Schnotz, W. (2002). Towards an integrated view of learning from text and visual displays. *Educational Psychology Review*, 14(1), 101-120.
- Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 49-69). New York: Cambridge University Press.
- Schnotz, W. & Baadte, C. (2008). Domain learning versus language learning with multimedia. In M. Farías & K. Obilinovic (Eds.), *Aprendizaje multimodal/ Multimodal learning* (pp. 21-49). Santiago de Chile: PUBLIFAHU USACH.
- Sökmen, A. (1997). Current trends in teaching second language vocabulary. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 237-257). Cambridge: Cambridge University Press.
- Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning and Technology*, 14(2), 50-73.
- Unsworth, L. (2006). Towards a metalanguage for multiliteracies education: Describing the meaning-making resources of language-image interaction. *English Teaching: Practice and Critique*, 5(1), 55-76.

Zimmerman, C. (1997). Historical trends in second language vocabulary instruction. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 5-19). Cambridge: Cambridge University Press.

## NOTES

- 1 Research for this study was financed by Project 031151OR from DICYT, Universidad de Santiago de Chile.
- 2 ANOVA value: 0.045 (sig)
- 3 ANOVA value: 0.045 (sig)