MODALITY OF INPUT AND VOCABULARY ACQUISITION

Tetyana Sydorenko
Michigan State University

This study examines the effect of input modality (video, audio, and captions, i.e., on-screen text in the same language as audio) on (a) the learning of written and aural word forms, (b) overall vocabulary gains, (c) attention to input, and (d) vocabulary learning strategies of beginning L2 learners. Twenty-six second-semester learners of Russian participated in this study. Group one (N = 8) saw video with audio and captions (VAC); group two (N = 9) saw video with audio (VA); group three (N = 9) saw video with captions (VC). All participants completed written and aural vocabulary tests and a final questionnaire.

The results indicate that groups with captions (VAC and VC) scored higher on written than on aural recognition of word forms, while the reverse applied to the VA group. The VAC group learned more word meanings than the VA group. Results from the questionnaire suggest that learners paid most attention to captions, followed by video and audio, and acquired most words by associating them with visual images. Pedagogical implications of this study are that captioned video tends to aid recognition of written word forms and the learning of word meaning, while non-captioned video tends to improve listening comprehension as it facilitates recognition of aural word forms.

INTRODUCTION

Input has received considerable attention in the field of second language acquisition (SLA) (Gass, 1997). In recent years, some interest has been expressed towards modality of input in language learning due to the increased use of multimedia materials. Multimedia, that is, a combination of print, audio, and imagery, has been argued to enhance input by making it more comprehensible (Plass & Jones, 2005). It has been shown that pictures and video can increase reading comprehension and listening comprehension (see Plass & Jones for a review). This supports Paivio's (1986, 1991, 2007) Dual Coding Theory, which states that a combination of imagery and verbal information improves information processing. The use of multimedia is also advocated because (a) it allows for the provision of authentic input and thus exposure to target culture, (b) it motivates learners, and (c) it accounts for students' different learning styles (Brinton, 2001). A considerable amount of research has also been conducted on the use of multimedia for vocabulary learning (see Plass & Jones for a review), but the findings are not entirely clear when captioned videos are used. Another area that deserves attention is how learners process different types of input presented simultaneously. While there is a theory of multimedia learning in the native language (Mayer, 1997, 2001), it has been applied to second language learning primarily in the context of multimedia annotations (Al-Seghayer, 2001; Jones, 2003; Jones & Plass, 2002; Plass, Chun, Mayer, & Leutner, 1998, 2003), but not in other multimedia environments. Given that vocabulary plays a key role in language acquisition and that videos are widely used as instructional material, this study investigates the effect of visual images, audio, and captions on a) the acquisition of vocabulary and b) learner attention to input.

LITERATURE REVIEW

Input from Visual and Auditory Modalities

Input in various modalities is now being used in language teaching because multiple modalities are believed to improve language acquisition. This view is supported by Paivio's (1986, 1991, 2007) Dual Coding Theory, the main assumption of which is that verbal and non-verbal stimuli are processed by two different systems, but that these verbal and non-verbal systems interact. The activation of both systems results in better recall. This explains why second language (L2) learning can be increased by combining

visual images with verbal information. Vocabulary learning from written text (Al-Seghayer, 2001; Chun & Plass, 1996a, 1996b; Plass et al., 1998, 2003) and aural passages (Jones & Plass, 2002) can be enhanced if new words are annotated with both verbal input and images rather than when they are annotated with only one of these stimuli. However, there are conflicting results on the effect of still pictures versus dynamic images. Al-Seghayer found that multimedia annotations consisting of video and text led to better vocabulary learning than annotations that combined pictures and text. However, the reverse was found by Chun and Plass (1996a). A possible reason for the different findings is the characteristics of the images, for example, their concreteness or familiarity. In studies with video input, gestures and facial expressions have been found to aid listening comprehension in the L2 (Hernandez, 2005; Sueyoshi & Hardison, 2005). However, Baltova (1994) argues that authentic videos help with global comprehension of information due to visual images, but they do not increase understanding of the language per se. To help language learners with comprehension of the language, videos are often augmented with on-screen text.

On-screen text can appear in various forms: subtitles (L1 text, L2 sound), reversed subtitles (L1 sound, L2 text), or captions (sound is in the same language as the text). Concerning comprehension, it is yet to be resolved which of these on-screen text presentations is most beneficial (Baltova, 1999; Lambert, Boehler, & Sidoti, 1981; Markham & Peter, 2003; Markham, Peter, & McCarthy, 2001). Danan (2004) suggests that subtitles should be used with very difficult material; otherwise, the use of captions in the L2 is advised. In vocabulary learning, L2 captions and reversed subtitles result in similar gains, and they are better than subtitles for recall (Baltova, 1999; Danan, 1992) and recognition (Lambert et al., 1981). This study focuses on captions in the L2 because captions appear to have an advantage over subtitles for vocabulary acquisition and they provide more exposure to the L2 than reversed subtitles.

Vocabulary Acquisition from Captioned Videos

While captions facilitate listening comprehension in a foreign language (Baltova, 1999; Garza, 1991; Guillory, 1998; Markham, 1993, 2001), their effect on vocabulary learning is not as transparent. Several studies investigated the influence of captions on learning vocabulary as assessed by written tests (Baltova, 1999; Danan, 1992; Neuman & Koskinen, 1992). At least one other study assessed the learning via aural tests (Markham, 1999). These researchers' tests, following Nation (2001), can be classified as recognition of form, recall of form (c-cloze, fill-in-the-blank, free recall), and recall of meaning (L2 to L1 translation). Table 1 summarizes these four studies.

Table 1. *Summary of Research on Vocabulary Acquisition from Captioned Videos*

Test	Study	Participants	Specific Task
Recognition			
• Written	Neuman & Koskinen, 1992	middle school ESL learners (beginning to advanced)	distinguish words from non-words
• Aural	Markham, 1999	college ESL learners (advanced)	select the word you hear
Recall			
• Written	Baltova, 1999	high school learners of French (beginning to intermediate)	c-cloze
	Danan, 1992	college learners of French (beginning; intermediate to advanced)	fill-in-the-blank, translation
	Neuman & Koskinen, 1992	middle school ESL learners (beginning to advanced)	free recall
• Aural	No studies		

Note. Estimated proficiency level is indicated in parentheses.

The studies cited above indicate that the presentation of video, audio, and captions (VAC) leads to better performance on both written and aural tests than the presentation of video and audio (VA), yet there are at least two gaps in the literature which make it difficult to generalize the findings. First, the studies have been conducted with learners of different ages and at different levels of proficiency. It is difficult to ascertain from the descriptions of the participants in the studies exactly how they would compare in terms of proficiency. Roughly, the participants in the studies can be categorized as beginning, intermediate, and advanced. Following this classification, it appears that aural vocabulary acquisition of beginning and intermediate learners has not been studied. Additionally, previous research has investigated the recognition, but not recall, of aural vocabulary. Form recognition is different from form recall or translation because less cognitive processing is required for form recognition (Nation, 2001). To determine which forms of vocabulary—written, aural, or both—are being learned from the VAC input, learner performance on recognition and recall vocabulary tests in written and aural modalities should be compared. Although acquisition of both written and aural forms of vocabulary is important in language learning, commonly used tests of vocabulary are based on orthography, not phonology (Milton & Hopkins, 2006). While it is generally assumed that learners can transfer their orthographic word knowledge to phonological word knowledge, this might not be the case. Previous research suggests that linguistic performance is best in the same modality in which new information was learned. This applies at least to semantic categorization in L1 (Dodd, Oerlemans, & Robinson, 1988) and vocabulary recognition in L1 (Nelson, Balass, & Perfetti, 2005) and L2 (Bird & Williams, 2002). Following this research, it is expected that after receiving aural input, learners will perform better on aural than on written vocabulary tests, and vice versa. However, it is necessary to investigate the outcome when both types of input are presented. Will the input in both aural and written modes be learned, and if so, at what ratio?

What Modalities do Learners Attend to?

It is important to investigate what learners pay attention to for both practical and theoretical reasons. While existing research indicates that the presentation of a video with audio and captions is superior at least for written vocabulary learning, the concern of many teachers is that learners might not attend to audio when they also have captions, which would hinder their listening skills development (Borras & Lafayette, 1994). Garza (1991) argues that captions help develop listening skills, but he did not investigate this claim empirically. Markham (1999) found that on an aural recognition test, performance of advanced learners was better when they watched captioned rather than non-captioned videos. He concluded that participants were attending to audio; however, it is not clear whether the same would be true for lower-level learners.

The hypothesized reason why learners might not be paying attention to audio when watching captioned video is that they have to divide their attention among three types of stimuli: visual images, text, and audio. Because their attentional capacity is limited, learners have to use attention selectively (Robinson, 2003; Wickens, 2007). Otherwise, learners' attempts to pay attention to all three modalities may result in cognitive overload. Cognitive overload occurs even when tasks are performed in the native language and is attributed to the limits of working memory (Baddeley, 1986, 1992; Chandler & Sweller, 1991; Miller, 1956; Sweller, 1999). The *redundancy principle* of the cognitive load theory (Sweller, 2005) is relevant for the current study. This principle assumes that redundant material slows down information processing and learning. Mayer (1997, 2001) applied cognitive load theory to the generative theory of multimedia learning. In one study, Mayer, Heiser, and Lonn (2001) found that native speakers of English who saw an animation and listened to a concurrent corresponding narration in their L1 were able to retain more information from the narration than those who also received captions as a third modality. The researchers concluded that captions are distracting when audio is also present because they carry the same information, which follows the redundancy principle. Thus, according to the cognitive load theory, non-captioned videos will be easier to process than captioned videos.

However, it appears that the prediction above is not borne out for second language learners. As the studies on captioned videos indicate, three modes of presentation (i.e., video, audio, and captions) are more beneficial for listening comprehension and vocabulary learning than two modes (i.e., video and audio). The question is, how do language learners as opposed to native speakers attend to the three types of

input? Do they continuously attend to audio, but switch between images and captions? Or could it be that language learners at times are not paying a lot of attention to visual images and instead focus on captions and audio, especially when the images do not carry useful information? It is also possible that language learners do not attend to audio as much as they do to captions or that they are not able to process audio as well as captions. For example, Lambert et al. (1981) found that on a combined written and aural test, L2 learners recognized more words learned from reading than those learned through listening. This might depend on the nature of instruction or input learners receive. If learners regularly receive more reading than listening practice, they might process captions better than audio. On the other hand, captions may be less beneficial for the learners whose L1 has a writing system that is very different from that of the L2. In the studies reviewed above, writing systems in the L1 and L2 are similar. In the current study with learners of Russian, the L1 and L2 orthographies differ, but learners of Russian at a college level master the Cyrillic alphabet within two weeks. Therefore, the difference in orthography is not a factor in this study.

To answer the question of whether learners attend to audio, one must compare the performance of the VAC (captioned video with audio) and the VC (captioned video without audio) groups. If both groups perform similarly post treatment, it would perhaps indicate that audio is not as necessary of a component for acquisition. However, if the VAC group outperforms the VC group, attention to audio could be deemed as beneficial to the learning process. If the VAC group underperforms the VC group, it might mean that attention was too split for the VAC group to succeed.

Only two studies have investigated what modalities L2 learners pay attention to when presented with three types of stimuli simultaneously: audio, video, and captions. In a study by Vanderplank (1988), European and Arabic high-intermediate to advanced college students learning English watched captioned British TV programs over the course of nine weeks. The European students were from France, Germany, Austria, Denmark, Italy, and Spain. Two learners reported that they tried but could not pay much attention to audio because the captions were present. Many students reported that initially they were distracted by the presence of captions. However, over the course of the study, European students found captions useful and not distracting, while the Arabic students mentioned that the captions changed too fast. Vanderplank suggested that European students are used to captions and can utilize them better. However, the difference between the L1 and L2 scripts might be the underlying factor (Winke, Gass, & Sydorenko, 2008). Another study was conducted by Taylor (2005) with beginning college learners of Spanish. Captions were found to be distracting for many learners with little exposure to Spanish, while more experienced learners could utilize all three types of stimuli when processing videos. The majority of all students (26 of 35) reported that they attended to the audio, suggesting that availability of captions does not make all students ignore the audio completely. Both Vanderplank and Taylor concluded that over time learners can develop strategies for processing input in the three modalities. While these studies provide useful information on learners' processing of captioned video, more studies are needed with a variety of learners and viewing conditions.

Research Questions

The major gaps in research on captioned video concern (a) how captions affect learners' acquisition of vocabulary, especially, the difference between written and aural word forms, (b) how exactly the learners are able to acquire vocabulary from videos, and (c) what input modalities learners pay attention to when watching videos. To fill this gap, the present study investigates the following four research questions.

1. Does the modality of input have a differential effect on the learning of written and aural forms of vocabulary?
2. Is the overall learning of words (i.e., combined written and aural vocabulary) affected by different input modalities?
3. What input modalities do the learners attend to when watching videos?
4. What strategies do learners use to acquire new vocabulary from videos?

The hypotheses are as follows.

1a. The VC group will score higher on written than on aural tests, and the VA group will score higher on aural than on written tests because studies suggest that performance is best in the modality in which new information was learned.

1b. The VAC group will score higher on written than on aural tests because Lambert et al. (1981) found that learners remembered more words from written than from aural input.

2a. The VAC group will perform better than the VA group based on the previous research on captioned video.

2b. The VC group will outperform the VA group since written input appears to have an advantage over aural input (Lambert et al., 1981).

2c. The VAC group will outperform the VC groups since there is some indication from previous research that learners pay attention to audio.

3a. Based on the existing studies, all learners will pay attention to captions.

3b. Learners will also pay attention to visuals because they have been shown to increase listening comprehension.

3c. Learners will pay less attention to audio than to captions because written input seems to be more beneficial than aural.

4. Learners will associate visual images with written and/or aural verbal information to learn new vocabulary, as has been shown to occur in studies on multimedia annotations.

METHODOLOGY

Participants

The participants were 26 non-heritage learners from two sections of second-semester (beginning) Russian at a large Midwestern university. The participants' age ranged from 18 to 26, with a mean of 20. The L1 was English for 25 participants and Cantonese for one. Of the native English speakers, one participant also considered French as an L1 and another considered Italian and Spanish as L1s in addition to English. There were 14 females and 12 males. The participants were compensated for their time with a gift certificate for lunch (valued at 5 dollars). They also received extra credit in their Russian course for participating.

The participants were divided randomly into three stimulus conditions: video with audio and captions (VAC), video with audio (VA), and video with captions (VC). There were eight participants in the VAC group, and nine participants in each of the other two groups. To ensure that the VAC, VA, and VC groups did not differ in terms of their abilities to learn written and aural word forms, these abilities were tested before the study.

Group Equality Test

All participants read one text and listened to another text two times, with the order and modality of texts counterbalanced (see Appendix A for one of the texts). The topics of the texts, professions and buying, had not been studied by the participants prior to the study. Almost every sentence in each text contained one new word accompanied by a visual image. None of the new words were cognates. Every sentence of the text, together with an accompanying picture when applicable, was presented on a PowerPoint slide for five seconds. The participants then took written and aural recognition and translation tests. On the recognition test, the participants had to mark the words that they thought appeared in the texts; these words were mixed with non-words. On the written recognition test, the new words from prior reading were presented on paper, and on the aural recognition test, the new words from prior listening were presented aurally. For the written and aural translation tests, the task was to translate from Russian into English the same target words that were on recognition tests. After that, the participants were asked to indicate for each target word whether they have (1) never encountered it before, (2) encountered it before, (3) knew its meaning, or (4) used it. Only those target words that were new for the participants were

included in the analysis. The results of a mixed-design ANOVA [Test (written recognition, aural recognition, written translation, aural translation) x Group (VAC, VA, VC)] showed a non-significant main effect of group, $F(2, 17) = .44$, $p = .65$, which means that the three treatment groups were comparable on learning new vocabulary. The results also revealed a significant main effect of test, $F(3, 51) = 18.42$, $p < .001$, $r = .51$ and a non-significant Test x Group interaction, $F(6, 51) = .96$, $p = .461$. As shown in Table 2 and Figure 1, the participants scored higher on written than on aural tests, suggesting that their reading abilities were better than their listening abilities.

Table 2. *Descriptive Statistics on the Group Equality Test*

	VAC (*N* = 8)		VA (*N* = 9)		VC (*N* = 9)		All Groups	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Written recognition	.84	.10	.83	.08	.87	.12	.85	.10
Aural recognition	.68	.11	.74	.11	.69	.19	.71	.14
Written translation	.62	.22	.69	.27	.57	.34	.62	.28
Aural translation	.27	.18	.27	.24	.48	.19	.35	.22

Note. The scores represent percentage of words learned from the new target words.

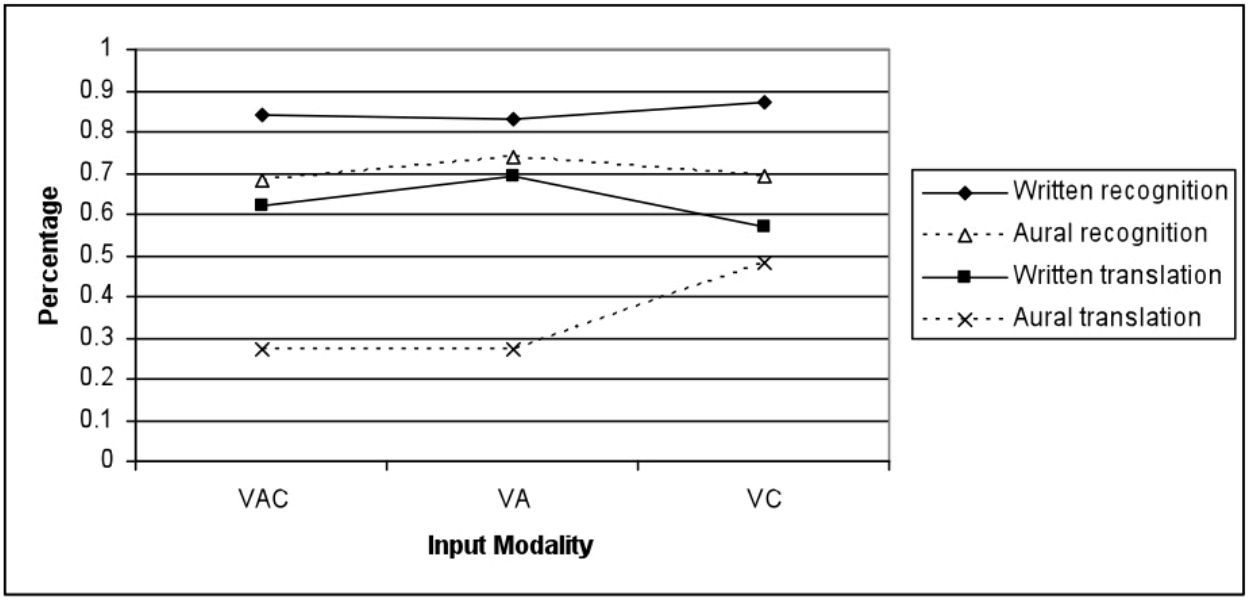

Figure 1. Results from the group equality test.

Procedures

The study took place in the language lab and was conducted outside of class time. The participants first filled out a background questionnaire, which consisted of questions in four categories: (a) demographic data (age, gender, L1); (b) length of study of L2 and motivation; (c) out-of-class exposure to L2; (d) self-rating of reading and listening skills in Russian. Then the participants watched the video clips on individual computers. The VAC group saw the videos with audio and captions, the VA group saw the videos with audio, and the VC group saw the videos with captions, but did not receive audio input. The participants were instructed to watch each video for meaning the first time, and to pay attention to the language the second time because they would be tested on both meaning and language. After watching each video, the participants completed comprehension, written and aural recognition, written and aural translation (from L2 to L1), and word knowledge tests in that order. The recognition test always preceded the translation test, while the order of written and aural tests was counterbalanced. Following Al-Seghayer

(2001), Bird and Williams (2002), and Danan (1992), the word knowledge test was not given before the study in order to avoid prompting the participants to pay special attention to new words. At the end of the study, the participants completed a final questionnaire. The whole procedure took between 50 and 60 minutes.

Materials

The materials consisted of three video clips, each 2 to 3 minutes long, from a popular Russian comedy series for native Russian speakers. The clips contained target words that were highly unlikely to be known since they did not appear in the participants' textbook. In addition, the participants' instructor felt that these words were highly unlikely to be heard in class. However, the participants could learn the target words from video context, mainly due to a high correlation between visual images and dialogs. For example, in one of the video clips a boy says *A u menya brat boksyor* ("My brother is a boxer"), after which a male boxer appears. That is, the criteria for target word selection were that words or phrases must be well-supported visually and most likely unknown to the participants.[1]

The topics of the video clips included professions, eating, and compliments. The participants knew only a few words on the topics of food and professions, and they had not studied the compliments topic prior to the study. There were 6 target words in Clip 1, 12 target words in Clip 2, and 10 target words in Clip 3 (see Appendix B). Of these 28 words, there were 13 nouns, 5 verbs, 5 adjectives, 1 adverb, 3 phrases consisting of 2 words, and 1 single-word expression. About one fourth of the words were abstract, others were concrete and well-supported visually. Of all target words, six were cognates, and two were false cognates. Usually, learners can guess new words from context when at least 95% of the words in the text are known (Nation, 2001). Following the intuition of the participants' Russian instructor, the videos in this study had a much lower percentage of known words.[2]

Instruments

Comprehension test

The comprehension test consisted of three very general true-false questions that did not require the knowledge of the target words. Although a larger number of questions is desirable to test comprehension, this was not possible because the videos were too short. The comprehension test was included to encourage the participants to pay attention not only to the language used in the videos, but also to the main ideas of the clips. Because comprehension of captioned videos has been widely studied, it was not the focus of this study and thus was not one of the dependent variables.

Recognition test

Since recognition of new lexical forms is considered to be an initial step in vocabulary learning, recognition of the target word forms was measured (see Appendix C). The recognition test involved discriminating between words presented in the input (target items) and those that were not presented in the input (non-target items) (Pulido, 2004). In this study, non-target items were non-words. Half of the target words and half of the non-words were presented in a written form, the other half aurally. Following Huibregtse, Admiraal, and Meara (2002), one point was awarded for a "yes" response to a target item, and one point was awarded for a "no" response to a non-word. Responses to non-words had to be taken into account to correct for guessing. Words that participants knew before the study were excluded from the analysis. The scores from the three videos were combined and calculated as the percentage of recognized new words from the total number of new words, following Smith (2004) (see Equation 1). Non-words were also added to the equation.

$$\frac{\text{recognized new words} + \text{unselected non-words}}{\text{all new words} + \text{all non-words}} \quad (1)$$

Translation Test

This test consisted of the same target words as the recognition test, but non-words were not included (see Appendix C). Half of the words were presented in a written form, the other half aurally. One point was given for each translation of the new word that was possible in the context of the video. For example, in clip 1 the word *lyotchik* means "pilot," but the translation *airplane* would have been accepted because in the clip the word *lyotchik* appeared when the airplanes were flying. Although this measure does not reflect the number of words the participants learned correctly, it shows some level of learning and the ability to remember new form-meaning associations. The scores from the three videos were combined and calculated as the percentage of new words translated from the total number of new words.

Word Knowledge Test

The purpose of the word knowledge test was to check which target words were new for each participant. The participants rated their knowledge of the words prior to the study as 1 = never encountered before, 2 = encountered before, 3 = know the meaning, 4 = use it (see Appendix C). Some non-words from the recognition tests were used to adjust for guessing. If a participant knew the meaning of the word prior to the study or used it, such a word was not considered new and was excluded from the analysis.

Final Questionnaire

In the final questionnaire, the participants were asked open-ended questions on whether they liked the videos and why, what strategies they used to learn the new words, what difficulties they had, and what could have helped them to understand the videos better. Two Likert-scale questions asked the participants to rate how much attention they paid to visual images and how helpful they were (following the design of Sueyoshi & Hardison, 2005). The participants were asked the same questions about audio and captions if they had them (see Appendix D).

Analysis

The between-participant independent variable was input modality (VAC, VA, and VC), and the within-participant independent variable was test modality (written and aural). There were two dependent variables: recognition and translation test scores. A mixed-design ANOVA[3] was conducted for each of the two dependent variables.

The data from open-ended questions on the final questionnaire were analyzed qualitatively, using a content analysis approach. In a content analysis approach, themes emerging from the data are grouped into categories, which should also emerge from the data (Berg, 2001). The data can then be subjected to frequency counts and descriptive statistics. The data from the two Likert-scale questions on the final questionnaire were analyzed using descriptive statistics. These data were not subjected to statistical tests due to the small number of tokens.

RESULTS

Does the modality of input have a differential effect on the learning of written and aural forms of vocabulary?

The results of a mixed-design ANOVA on the vocabulary recognition test [Test (written, aural) x Input Modality (VAC, VA, VC)] showed a non-significant main effect of test, $F(1, 23) = 1.32$, $p = .263$. The results also revealed a significant Test x Input Modality interaction, $F(2, 23) = 4.06$, $p = .031$, $r = .39$. Groups with captions scored higher on written than on aural recognition, while the VA group scored higher on aural than on written recognition (see Table 3 and Figure 2). The results of a mixed-design ANOVA on the vocabulary translation test [Test (written, aural) x Input Modality (VAC, VA, VC)] showed a non-significant main effect of test, $F(1, 23) = 2.43$, $p = .133$, and a non-significant Test x Input Modality interaction, $F(2, 23) = .53$, $p = .599$. This indicates that there were no significant differences between written and aural translation for each group.

Is the overall learning of words (i.e., combined written and aural vocabulary) affected by different input modalities?

The results of the mixed-design ANOVA on the vocabulary recognition test (mentioned above) showed a non-significant main effect of input modality, $F(2, 23) = 1.10$, $p = .351$. This means that combined written and aural recognition was the same for each group. The results of the mixed-design ANOVA on the vocabulary translation test (described above) showed a significant main effect of input modality, $F(2, 23) = 3.75$, $p = .039$, $r = .37$. A post-hoc Tukey's HSD test revealed a significant difference between the VAC and the VA groups. The VAC group scored significantly higher than the VA group and non-significantly higher than the VC group on overall translation. This suggests that the VAC combination is more favorable than the VA combination for learning the meanings of new words.

Table 3. *Descriptive Statistics on Vocabulary Tests*

	VAC ($N = 8$)		VA ($N = 9$)		VC ($N = 9$)		All Groups	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Written recognition	.73	.10	.63	.10	.76	.12	.71	.12
Aural recognition	.67	.15	.69	.08	.68	.06	.68	.10
Written translation	.36	.11	.25	.13	.28	.10	.30	.12
Aural translation	.35	.18	.18	.13	.24	.12	.25	.15

Note. The scores represent percentage of words learned from the new target words.

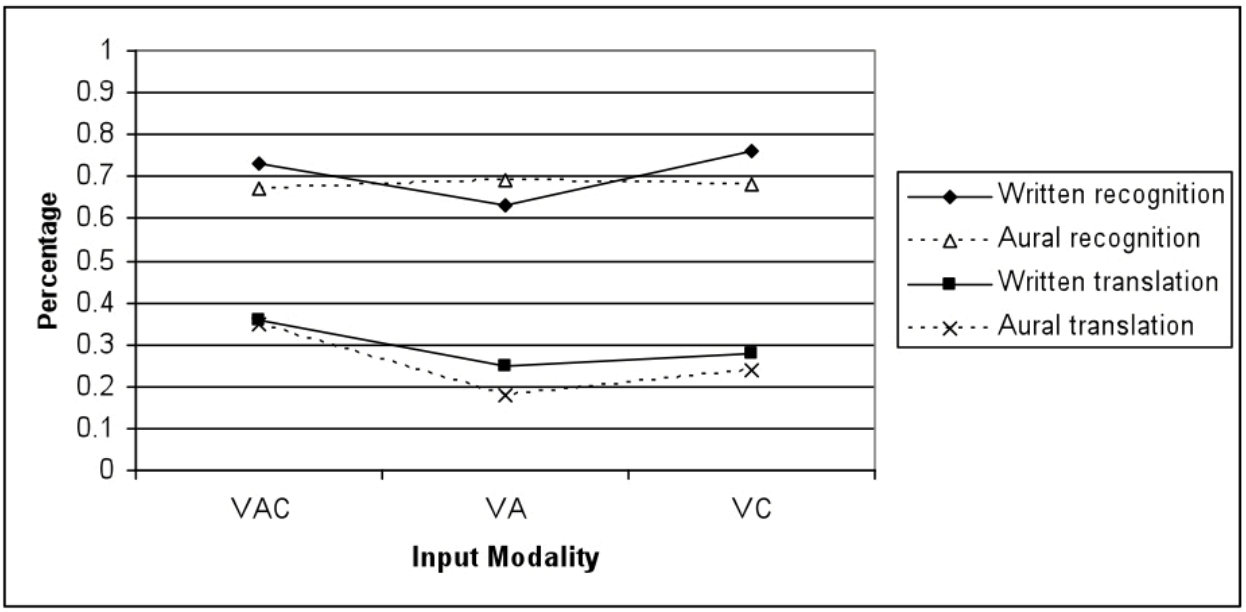

Figure 2. Results from vocabulary tests.

What input modalities do the learners attend to when watching videos?

To answer this research question, the participants were instructed to rate the amount of attention they paid to video, audio, and captions, as well as the usefulness of these modalities. They were also asked to describe the difficulties they encountered when watching the videos and taking vocabulary tests, as well as how these difficulties can be avoided.

The descriptive statistics for attention paid to various modalities are reported in Table 4. In the VAC group, the participants thought they paid most attention to captions, then to video, then to audio. In the

VA group, the participants said they paid the same amount of attention to audio and video. In the VC group, the participants reported paying more attention to captions than to video. That is, both the VAC and the VC groups seemed to pay most attention to captions.

Table 4. *Participants' Perception of the Amount of Attention Paid to Audio, Captions, and Video*

	VAC (*N* = 8)		VA (*N* = 9)		VC (*N* = 9)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Audio	3.75	1.28	4.33	0.71		
Captions	4.5	1.07			4.11	0.78
Video	4.25	0.71	4.33	0.71	3.56	0.88

Note. Higher means indicate more reported attention.

The participants were also asked to rate the utility of video, as well as captions and audio when available, for their understanding of the clips. The descriptive statistics from these data are reported in Table 5. In all three groups, the participants found video to be most helpful, although, as mentioned above, none of the three groups reported paying most attention to video. The VAC group found audio to be the least helpful.

Table 5. *Participants' Perception of the Amount of Help Obtained from Audio, Captions, and Video*

	VAC (*N* = 8)		VA (*N* = 9)		VC (*N* = 9)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Audio	3.25	1.39	3.11	0.93		
Captions	4.13	0.99			4.00	1.00
Video	4.75	0.46	4.78	0.44	4.67	0.71

Note. Higher means indicate more reported helpfulness.

Reported difficulties with watching the videos and completing vocabulary tests and their solutions also reveal to what modalities the learners were attending (see Table 6). Participants in all groups had some difficulty with audio or captions. Such difficulties were contributed to the speed of the dialogs (they were too fast), lack of time to read all captions, and the heavy burden of reading captions while watching the videos at the same time. One learner also mentioned focusing only on known words due to captions. Another learner reported reading captions and scanning the images, but having no time to listen to audio. One more learner tried to sound out captions because there was no audio. Because learners in the VAC group did not specify whether a "fast dialog" and "very fluent Russian" referred to audio, captions, or both, it is not clear which kind of stimulus as a whole was more difficult to process, captions or audio.

Participants in all groups reported that they had difficulty with unknown vocabulary and vocabulary tests. Many participants mentioned that they could not figure out the meanings of new words or that there was too much new vocabulary. One participant pointed out that it was difficult to learn words not supported by images. Learners also mentioned not remembering which specific words were used in the videos, especially their aural forms. Some participants reported that they could guess the meanings of the words while watching the videos, but forgot the actual words by the time they had to take a vocabulary test.

Table 6. *Difficulties with Watching Videos and Completing Vocabulary Tests and their Mitigation*

	VAC (*N* = 8)	VA (*N* = 9)	VC (*N* = 9)
Difficulties			
With audio or captions			
fast dialog	2	2	
fast flow of captions			5
very fluent Russian	1		
trying to read captions and watch videos at once			1
captions make one focus on known words	1		
only read captions and scanned visual images	1		
trying to sound out captions due to lack of audio			1
With vocabulary in the videos and tests			
too much new vocabulary	1	4	
figuring out meanings of new words	3	3	4
learning words not supported by images		1	
were in the video	2	3	3
remembering aural word forms from the video	1		2
translating	1	3	6
remembering word-meaning associations	5	3	
What can lessen the difficulties?			
Scaffolding			
shorter sentences in the dialog	1		
captions		3	
slower dialog		2	
sound			7
more time for reading captions			1
Knowledge of Russian			
more knowledge of Russian	3		
more known or pre-taught vocabulary		5	2
more time to get familiar with pre-viewing vocabulary	1		
dictionary	1		

Note. The numbers indicate how many within the group provided the given response.

The learners were also asked what could have helped them understand the videos better. The participants mentioned various types of scaffolding (depending on the group they were in), such as shorter sentences, captions, sound, more time to read captions, or slower dialog. It appears that most learners wanted to have access to all modalities (video, audio, and captions) and have more time to process the information in these modalities. However, while only two learners in the VC group did not state that they would like to

have audio, six learners in the VA group did not say anything about captions. That is, it is either more important or more natural for the learners to watch videos with audio than with captions, as one would expect. Several learners provided another solution to the problem of not understanding the videos: a better knowledge of Russian in general or vocabulary in particular.

What strategies do learners use to acquire new vocabulary from videos?

On the final questionnaire, the participants were asked whether they learned new words from the videos and how they were able to do that. The strategies learners reported and their frequencies are provided in Table 7 and can be divided into two categories: modality-specific strategies and common vocabulary guessing strategies.

Regarding modality-specific strategies, six learners from each group reported using visual images to help them figure out the meanings of new words. One participant wrote, "Most words I learned were accompanied by actions on screen, such as *sadites'* ["sit down"], *proshu vas* ["after you"], and *boksyor* ["boxer"]." The participants in the VAC group did not say whether they matched visual images to captions, audio, or both.

Concerning common vocabulary guessing strategies, only one participant in the VAC group reported using them; specifically, this learner relied on familiar roots. More participants in the other two groups mentioned that they used guessing strategies: five participants in the VA group and four in the VC group. The participants thought they had understood new words which were similar to their L1s (English, Spanish, or Italian), although they did not realize that some of these words were false cognates. They also reported using roots of familiar words, relying on the verbal context, and employing their knowledge of grammar to understand new words.

Table 7. *Learners' Strategies and their Frequencies for Learning New Words*

Strategies	VAC (*N* = 8)	VA (*N* = 9)	VC (*N* = 9)
Modality-specific strategies			
matching visual images with words	6	6	6
reading captions	1		
Common vocabulary guessing strategies			
recognizing words that are similar to L1		3	1
using the roots of known words	1		1
paying attention to verbal context		1	1
paying attention to grammar		1	1

Note. The numbers indicate how many within the group provided the given response.

DISCUSSION

The intent of this study was to investigate vocabulary acquisition from different types of video input when the goal is to both understand the content of the videos and to learn new words from them. The results suggest that for beginning learners with better reading than listening skills: (a) captions facilitate recognition of written word forms, while audio facilitates recognition of aural word forms; (b) more word meanings are learned when videos are shown with both audio and captions than with either audio or captions; (c) participants think they pay most attention to captions, then to video, then to audio, but they consider video to be the most helpful; some participants have difficulty attending to all three modalities; and (d) the meanings of some new words can be learned from very difficult authentic videos when the language is well-supported by visual images. These findings are discussed below in detail.

Does the modality of input have a differential effect on the learning of written and aural forms of vocabulary?

Since the VA group performed better on aural than on written recognition test, and the performance of the VC group resulted in the reverse pattern, the results support the hypothesis that recognition of form is best when modality of input and test modality are the same, as was found by Bird and Williams (2002). Jones (2004), on the other hand, found that vocabulary recognition was not affected by the modality of the test. However, in her study text was in L1, while in this study captions and vocabulary tests were in L2. Additionally, in Jones' study test modalities were written and pictorial, while in this study they were written and aural.

Contrary to the hypothesis, it was found that for recall of meaning test modality does not interact with input modality. As mentioned earlier, recognition and translation of vocabulary are different skills (Nation, 2001). The recognition test indicates whether learners have noticed the forms in the input, that is, their episodic memory of the forms, while the translation test indicates whether learners have understood the meaning of the forms (Pulido, 2004). Compared to form recognition, production of meaning requires deeper processing because learners need to deduce the meaning of the form while they watch the videos, and then recall the meaning of the form when they take the test. It is possible that if learners understood and remembered the meaning of the word, they have built the connections between the meaning and both of its forms (written and aural). Thus it does not matter whether they have to produce the meaning of the written or of the aural form of the same word, or at least the differences are not substantial. On the other hand, if learners have only noticed the form either through reading or listening, but did not understand its meaning, one can suppose that they have not built connections between the written and aural forms.

For form recognition, the results also support the hypothesis that given both written and aural input, learners presented with video, audio, and captions would perform better on written than on aural tests. This could be due to instruction because the participants in this study appeared to have better reading than listening skills, or due to the previous finding that people in general process written input better than aural (Nelson et al., 2005) at least when non-logographic script is used.

Is the overall learning of words (i.e., combined written and aural vocabulary) affected by different input modalities?

The hypothesis was that learners would be able to recognize and translate more vocabulary in the VAC group than in the VA group. That is, captions were predicted to increase the acquisition of vocabulary. This was only partially confirmed: while the VAC group scored significantly higher than the VA group on overall translation, there were no differences between the groups on overall recognition. It was mentioned earlier that recognition of written and aural forms depends on modality of input, so there is an interaction effect. It appears that for acquiring word meaning, the most beneficial condition is video combined with both written and aural verbal input, but for form recognition, there is no effect from combined written and aural verbal input.

The fact that learners in the VAC group were able to acquire more word meanings than those in the VA group is in line with findings in previous research on captions (Baltova, 1999; Danan, 1992; Neuman & Koskinen, 1992). However, this runs counter to Robinson's (2003) hypothesis that attentional division among many stimuli may be negatively taxing for L2 learners. Even though the learners in the VAC group had to divide their attention among more stimuli than those in the VC and the VA groups, and even reported that following audio or captions was cognitively taxing, the process of doing so did not hinder their language learning. In fact, in this study it was found that performing three tasks was better than performing two. This finding also opposes the predictions of cognitive load theory: according to the redundancy principle, the presentation of the same information simultaneously through two modalities (text and audio) negatively affects information processing at least in L1 (Sweller, 2005). One explanation why native speakers and L2 learners may process information in multiple modalities differently is the idea that certain kinds of redundancy, such as repetitions, topic-fronting, or paraphrase, can be beneficial for learners' comprehension of input (Larsen-Freeman & Long, 1991). In multimedia environments, redundancy stemming from different modalities might be also beneficial. Jones (2003) and Grgurović and

Hegelheimer (2007) attribute the advantage of redundancy through multiple modalities to learners' individual preferences: learners choose what they need to focus on (video, audio, or captions). Additionally, while the processing of speech is natural for native speakers, learners, especially at the beginning level, might not even know where each word begins and ends. As Ellis (2003, p. 77) put it, "learning to understand a language involves parsing the speech stream into chunks which reliably mark meaning." Captions could be more useful than distracting because they help learners parse the stream of speech, as was suggested by Vanderplank (1993) and reflected in the learners' comments in Winke et al. (2008).

It is also possible that captions help beginners more than audio in learning new vocabulary from videos because the VC group scored higher than the VA group. However, because this difference was not significant, further research is needed. On the other hand, even if learners can process captions better than audio, there is evidence in this dataset that they still attend to audio because the VAC group was able to recall more word meanings than the VC group, although this difference was not significant. In other words, attention (at some level) was paid to audio and can be considered as the factor that increased word recall. This supports Markham's (1999) finding that learners attend to audio when they also have captions.

Although the VC group slightly outperformed the VA group, it is not suggested that learners should be exposed to the VC rather than the VA input. The VC group was used only for research purposes and is generally not suitable for instruction because, as evident from the participants' comments, this condition is unnatural. One learner tried to sound out captions because there was no audio, and most learners in the VC group said they wanted audio.

What do the learners attend to when watching videos?

The participants reported paying most attention to captions, then video, then audio. Thus, as predicted, learners paid attention to all three modalities, and they paid more attention to captions than to audio. This suggests that captions might have an advantage over audio, which goes along with the results of the vocabulary tests: in the VAC group written recognition was higher than aural, and the VC group outperformed the VA group on the translation test. The participants also reported paying some attention to audio rather then completely ignoring it, although one participant in the VAC group did mention that he/she ignored the audio while trying to process video and captions. It is not surprising that participants paid attention to video because studies on listening comprehension (Hernandez, 2005; Sueyoshi & Hardison, 2005) and multimedia annotations (Al-Seghayer, 2001; Chun & Plass, 1996a) found that video enhances comprehension and vocabulary learning as compared to verbal information. Wagner (2007) found that when listening to a lecture or a dialog, all learners paid attention to video, although to varying degrees.

While learners seem to attend more to captions than to video, the participants reported that video is more helpful than captions. A plausible explanation is that language learners can only understand a portion of captions in L2, but they have no difficulty processing visual images and thus find them more helpful. This goes along with the fact that learners in all groups were reporting difficulties following audio or captions, but not video, which supports previous findings (Taylor, 2005; Vanderplank, 1988). In Jones (2003), some learners reported that although annotations in the form of L1 translations were helpful, they did not encourage deep thinking. On the other hand, pictures did lead to deep processing of aural input, especially when combined with text. If learners find L1 translations less helpful than images, they will surely find captions in L2 less helpful than video. Studies reviewed in Paivio (2007) support learners' reports that imagery has an advantage over verbal input for information processing.

An interesting finding was that many learners preferred to have access to all modalities (video, audio, and captions), even though they also reported difficulties processing all three modalities. While this seems to be a contradiction, it could be that some cognitively taxing tasks are beneficial for language learners, especially when multimedia is involved. In Jones (2003), learners who did not have access to annotations during a listening task were frustrated because they could not comprehend the input, while learners who had annotations in various modalities did not express such concerns. Thus, for language learners a combination of input in multiple modalities may be more of an advantage rather than a distraction.

What strategies do learners use to acquire new vocabulary from videos?

As predicted, it appears that visual context in combination with either aural or written language is a frequently used strategy to learn new words from videos. The same was found in previous research on multimedia annotations (Al-Seghayer, 2001; Chun & Plass, 1996a; Jones & Plass, 2002; Plass et al., 1998, 2003). The detailed analysis of the translation test results shows that the clarity of visuals is important: on average, 36% of words well-supported by visual images were translated correctly, but only 6% of concrete and abstract words which were not clearly represented by images were translated correctly. One participant also pointed out that it was difficult to learn words not supported by images. However, readers should keep in mind that visual images often matched the language in these videos, which is not necessarily the case in all videos. In terms of the differences between groups, participants in the VAC group reported using fewer general guessing strategies than the participants in the other two groups. This could be because they were paying attention to all three modalities and did not have time to employ other strategies, or because visual images in conjunction with audio and captions were enough to understand the meanings of many new words. These findings relate especially to videos that were difficult for the learners. It is possible that learners at higher proficiency levels would have used more general guessing strategies.

It should be noted that although all learners acquired some vocabulary (between 3 and 7 new words), based on the comments from the final questionnaire, many participants did not feel that they learned new words. The absence of feedback may explain these results. The participants often guessed the meanings of the words, but they did not know whether they learned them correctly. In fact, 19% of all accepted translations were incorrect, although possible in the context of the videos. In addition, two learners said "I don't know that I learned any new words to the extent that I would feel comfortable using them—but I could probably recognize a few" and "I comprehended cognates, but cannot use them." Additionally, several participants reported that they could guess the meanings of the words while watching the videos, but forgot the actual words by the time they had to take a vocabulary test. To convince the students that they can learn some vocabulary from videos, teachers should provide learners with feedback and opportunities for production.

CONCLUSION

The pedagogical implication of this study is that different types of video input seem to provide different benefits. First, teachers can use videos to improve their students' vocabulary knowledge. Vocabulary knowledge is often a big hurdle for language learners, and it seems that captioned video with content well-supported by visual images can be a useful type of input, especially for beginning learners who often receive little authentic input because it is deemed as too difficult. However, when the goal is to improve listening, another skill with which foreign language learners often have difficulties, videos should be without captions. Teachers can also show a video once with captions and once without if the goals are both vocabulary learning and listening skills development. Videos can be used not only for learning new vocabulary, but also to reinforce initially learned vocabulary because the combination of images and verbal forms in the aural or written mode helps subsequent recall of vocabulary (Mayer, 2001; Paivio, 2007). Finally, learners should be exposed to captioned videos so that they can develop strategies for dealing with several types of input at the same time (Vanderplank, 1988).

Several questions were answered in this study, but future research is necessary to continue investigating how L2 learners process captioned videos. The findings of this study are limited in several ways. First, a larger sample size would increase power, that is, the probability that a test will find an effect assuming that it exists in the population. Second, a better control of target word selection, such as exclusion of cognates, would have been desirable, but this was constrained by the availability of videos in which new words were well-supported visually. The extreme shortness of vocabulary tests was also due to the available videos, which affected the test reliability and thus the validity of the findings. It is advisable to replicate this study with a larger number of videos and longer vocabulary tests. If such videos are not available, the researchers could consider creating their own videos. Another methodological issue is the way learners' strategies and their viewing behavior were investigated. While learners' reports of what

they paid attention to provide directions for further research, these results should be interpreted with caution because no statistical analyses were performed, and the reliability of self-reported data is not clear. Additionally, on the final questionnaire the participants reported some of the strategies they employed for learning the new words from videos, but their explanations were not very detailed, possibly because learners themselves are not aware of this process. Verbal reports, such as stimulated recall or think-aloud protocols, might be a better method for investigating the processing of video. Eye-tracking technology could be also used to investigate how learners attend to video and captions, but not audio.

The generalizability of the findings is also restricted by the nature of the videos and participant pool. The videos were of the comedy genre rather than informative, and they contained a large number of vivid images, which was conducive to the learning of new vocabulary. If videos had a heavier information load, for example, depicting historic events, the learners could have paid more attention to the propositional content rather than new vocabulary. The participants in this study were beginners at a university foreign language program in the US. They did not often watch videos with captions in Russian, at least as part of their course, and thus might have not developed strategies for dealing with different types of input simultaneously. These learners also appeared to have better reading than listening skills. Thus, the results cannot be generalized to learners of different proficiency levels, in different contexts, such as the target language environment, or with different types of instruction. Finally, script differences can play a role in the way learners process captions (Winke et al., 2008). Native speakers of English learning Chinese or Arabic could receive fewer benefits from captions in vocabulary learning than learners of Russian or German, or their processing strategies might be different.

Future studies could take different directions to investigate vocabulary acquisition from captioned videos. One factor not taken into account in this study is individual differences in modality preferences, which could influence vocabulary acquisition from videos. Dörnyei (2005) suggested that some people are visual and others are auditory learners. Spatial and verbal working memory might also contribute to individual differences affecting cognitive processing of videos since this was a factor in studies on multimedia annotations (Plass et al., 1998, 2003). The researchers could also look into other beneficial ways of using captions in language instruction. For example, as one anonymous reviewer suggested, captions in a form of dynamic glosses are now possible with recent technological developments and should be investigated. Giving learners some control over their learning can increase the benefits of instruction. In Jones (2003), participants liked the fact that they could choose from multiple modalities, and in this study learners wished they had control over the number of times they could play the videos. Finally, more research is needed to understand how language learners as opposed to native speakers learn in multimedia environments as the results in this study suggest that the processes might be different.

NOTES

1. The initial criteria for target word selection were not strict, which made it possible to identify the maximum number of possible target words. This was necessary in order to maximize the length of the vocabulary tests, thus making them more reliable. For example, even though many participants might have known the word *holodnij* (“cold”), it was nevertheless included in the vocabulary test. Ideally, videos with a large number of unknown target words that were well-supported visually would be used, but I was not able to find such videos. I was also not able to find more videos of this kind. The solution to this limitation was the use of the word knowledge test to identify which words were new for each participant. In addition, the target words were counterbalanced across written and aural vocabulary tests in such a way that half of the words predicted to be possibly known were on the aural test, and the other half on the written test. For the same reason, words that are easy to learn, such as cognates, were included. However, they were evenly divided between written and aural vocabulary tests.

2. While a more principled approach of measuring the ratio of known/unknown words in the videos would have been to give the participants a word knowledge test of all words in the videos, this was not done because the participants were willing to volunteer for a limited amount of time. The use of the instructor’s judgments was considered acceptable because the participants were in a foreign language

environment, and only three participants had completed a study-abroad program (as they indicated in the background questionnaire). That is, most students had limited exposure to vocabulary not used in the classroom. Since the ratio of known/unknown words in the videos was used to describe the nature of the videos rather than as a variable, a more stringent measure was not considered crucial for this study.

3. Although the sample size in each group was relatively small, the homogeneity of variances assumption was met, thus it was appropriate to use an ANOVA test.

ACKNOWLEDGMENTS

I would like to thank Dr. Paula Winke, Dr. Susan Gass, Dr. Shawn Loewen, Sara Hillman, and Maren Schierloh for reviewing earlier drafts of this report. I am also thankful to Drs. Diana Pulido and Senta Goertler for valuable suggestions. Finally, I would like to thank Dr. Dennie Hoopingarner and Michael Kramizeh for technical support. All remaining errors are my own.

ABOUT THE AUTHOR

Tetyana Sydorenko is a Ph.D. candidate in the Second Language Studies Program at Michigan State University. Her research interests include computer-assisted language instruction and testing, second language curriculum and materials development, and second language acquisition. She has published book chapters on computer-assisted language learning and testing.

E-mail: sydoren1@msu.edu

REFERENCES

Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology, 5*(1), 202–232. Retrieved from http://llt.msu.edu/vol5num1/alseghayer/default.html

Baltova, I. (1994). The impact of video on the comprehension skills of core French students. *The Canadian Modern Language Review, 50*, 507–532.

Baltova, I. (1999). Multisensory language teaching in a multidimensional curriculum: The use of authentic bimodal video in core French. *The Canadian Modern Language Review, 56*(1), 32–48.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D. (1992). Working memory. *Science, 255*, 556–559.

Berg, B. L. (2001). An introduction to content analysis. In B. L. Berg (Ed.), *Qualitative research methods for the social sciences* (pp. 238–267). Boston: Allyn and Bacon.

Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics, 23*, 509–533.

Borras, I., & Lafayette, R. C. (1994). Effect of multimedia courseware subtitling on the speaking performance of college students of French. *The Modern Language Journal, 78*, 61–75.

Brinton, D. (2001). The use of media in language teaching. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (pp. 459–476). Boston, MA: Heinle & Heinle.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293–332.

Chun, D. M., & Plass, J. L. (1996a). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal, 80,* 183–197.

Chun, D. M., & Plass, J. L. (1996b). Facilitating reading comprehension with multimedia. *System, 24,* 503–519.

Danan, M. (1992). Reversed subtitling and dual coding theory: New directions for foreign language instruction. *Language Learning, 42*(4), 497–527.

Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta, 49*(1), 67–77.

Dodd, B., Oerlemans, M., & Robinson, R. (1988). Cross-modal effects in repetition priming: A comparison of lipread, graphic and heard stimuli. *Visible Language, 22,* 58–77.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Erlbaum.

Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 63–103). Malden, MA: Blackwell.

Garza, T. J. (1991). Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign Language Annals, 24*(3), 239–258.

Gass, S. M. (1997). *Input, interaction, and the second language learner*. Mahwah, NJ: Erlbaum.

Grgurović, M., & Hegelheimer, V. (2007). Help options and multimedia listening: Students' use of subtitles and the transcript. *Language Learning & Technology, 11*(1), 45–66. Retrieved from http://llt.msu.edu/vol11num1/pdf/grgurovic.pdf

Guillory, H. G. (1998). The effects of keyword captions to authentic French video on learner comprehension. *CALICO Journal, 15*(1-3), 89–108.

Hernandez, S. S. (2005). The effects of video and captioned text and the influence of verbal and spatial abilities on second language listening comprehension in a multimedia learning environment. *Dissertation Abstracts International, 65*(8), 2958-A-2959-A. (UMI No. DA3142667)

Huibregtse, I., Admiraal, W., & Meara, P. (2002). Scores on a yes-no vocabulary test: Correction for guessing and response style. *Language Testing, 19,* 227–245.

Jones, L. (2003). Supporting listening comprehension and vocabulary acquisition with multimedia annotations: The students' voice. *CALICO Journal, 21*(1), 41–65.

Jones, L. (2004). Testing L2 vocabulary recognition and recall using pictorial and written test items. *Language Learning & Technology, 8*(3), 122–143. Retrieved from http://llt.msu.edu/vol8num3/jones/default.html

Jones, L., & Plass, J. L. (2002). Supporting listening comprehension and vocabulary acquisition with multimedia annotations. *The Modern Language Journal, 86,* 546–561.

Lambert, W. E., Boehler, I., & Sidoti, N. (1981). Choosing the languages of subtitles and spoken dialogues for media presentations: Implications for second language education. *Applied Psycholinguistics, 2,* 133–148.

Larsen-Freeman, D., & Long, M. (1991). *An introduction to second language acquisition research*. London: Longman.

Markham, P. (1993). Captioned television videotapes: Effects of visual support on second language comprehension. *Journal of Educational Technology Systems, 21*(3), 183–191.

Markham, P. (1999). Captioned videotapes and second-language listening word recognition. *Foreign Language Annals, 32*(3), 321–328.

Markham, P. (2001). The influence of culture-specific background knowledge and captions on second language comprehension. *Journal of Educational Technology Systems, 29*(4), 331–343.

Markham, P., & Peter, L. (2003). The influence of English language and Spanish language captions on foreign language listening/reading comprehension. *Journal of Educational Technology Systems, 31*(3), 331–341.

Markham, P., Peter, L. A., & McCarthy, T. J. (2001). The effects of native language vs. target language captions on foreign language students' DVD video comprehension. *Foreign Language Annals, 34*(5), 439–445.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*, 1–19.

Mayer, R. E. (2001). *Multimedia learning.* New York: Cambridge University Press.

Mayer, R. E., 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.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.

Milton, J., & Hopkins, N. (2006). Comparing phonological and orthographic vocabulary size: Do vocabulary tests underestimate the knowledge of some learners? *The Canadian Modern Language Review, 63*(1), 127–147.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. New York: Cambridge University Press.

Nelson, J., Balass, M., Perfetti, C. (2005). Differences between written and spoken input in learning new words. *Written Language and Literacy, 8*(2), 101–120.

Neuman, S. B., & Koskinen, P. (1992). Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. *Reading Research Quarterly, 27*, 94–106.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford: Oxford University Press.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology, 45*, 255–287.

Paivio, A. (2007). *Mind and its evolution: A dual coding theoretical approach*. Mahwah, NJ: Erlbaum.

Plass, J. L., Chun, D. M., Mayer, R. E., & 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. L., Chun, D. M., Mayer, R. E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior, 19*, 221–243.

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.

Pulido, D. (2004). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning, 54*(3), 469–523.

Robinson, P. (2003). Attention and memory during SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 631–678). Malden, MA: Blackwell.

Smith, B. (2004). Computer-mediated negotiated interaction and lexical acquisition. *Studies in Second Language Acquisition, 26*, 365–398.

Sueyoshi, A., & Hardison, D. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning, 55*(4), 661–699.

Sweller, J. (1999). *Instructional design in technical areas.* Camberwell, Australia: ACER Press.

Sweller, J. (2005). The redundancy principle in multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 159–168). New York: Cambridge University Press.

Taylor, G. (2005). Perceived processing strategies of students watching captioned video. *Foreign Language Annals, 38*(3), 422–427.

Vanderplank, R. (1988). The value of teletext sub-titles in language learning. *English Language Teaching Journal, 42*(4), 272–281.

Vanderplank, R. (1993). A very verbal medium: Language learning through closed captions. *TESOL Journal, 3*(1), 10–14.

Wagner, E. (2007). Are they watching? Test-taker viewing behavior during an L2 video listening test. *Language Learning & Technology, 11*(1), 67–86. Retrieved from http://llt.msu.edu/vol11num1/wagner/default.html

Wickens, C. D. (2007). Attention to the second language. *IRAL, 45*(3), 177–191.

Winke, P., Gass, S., & Sydorenko, T. (2008, August). *The effects of captioning on video-based listening activities in the second language classroom.* Paper presented at the International Association of Applied Linguistics conference, Essen, Germany.

APPENDIX A. One of Two Group Equality Tests

Лена: Какие ты купила подарки на Новый Год?
Lena: What presents did you buy for the New Year?

Оля: Сестре и брату я купила игрушки.
Olya: For my sister and my brother, I bought toys.

Оля: Маме я купила чулки.
Olya: For my mom, I bought panty hose.

Оля: Тёте—платок.
Olya: For my aunt, I bought a head scarf.

Оля: Папе я купила ремень.
Olya: For my dad, I bought a belt.

Оля: А ты что купила?
Olya: And what did you buy?

Лена: А я папе купила перчатки.
Lena: For my dad, I bought gloves.

Лена: Маме я купила шляпу.
Lena: For my mom, I bought a hat.

Лена: Сестре—шкатулку.
Lena: For my sister I bought a jewelry box.

Лена: А брату я дам деньги.
Lena: And my brother will get money.

Note. Target words supported by visual images are underlined here, but not in the study. Translation is provided for the journal readers only, not for the participants.

APPENDIX B. Target Words from Videos

Video 1	Video 2	Video 3
boksyor (“boxer”)	*yesh* (“eat”)	*sadites’* (“sit down”)
ohotnik (“hunter”)	*ne budu* (“I won’t”)	*tsvetochki* (“flowers”)
iz OMONa (“from SWAT”)	*morkovka* (“carrot”)	*obaldet’* (“wow”)
tankist (“tank operator”)	*sol’* (“salt”)	*divnij* (“nice”)
lyotchik (“pilot”)	*lapsha* (“noodles”)	*proshu vas* (“after you”)
komandir korablya (“ship commander”)	*lupa* (“magnifying glass”)	*pohozhi* (“look like”)
	goryachij (“hot”)	*naryadnaya* (“dressed up”)
	holodnij (“cold”)	*prichyoska* (“hairdo”)
	lopaj (colloquial “eat”)	*potresayushche* (“super”)
	luk (“onion)	*kulonchik* (“pendant”)
	kompot (“compote”)	
	korotkaya (“short”)	

Note. All captions (including target words) appeared in Cyrillic in the study.

APPENDIX C. Sample Comprehension, Vocabulary, and Word Knowledge Tests for Videos

Mark the following as true or false.

____ The boys' friends will help them fight.

____ The boys are saying they will be in the military when they grow up.

____ The boys are brothers.

Check all words/phrases that were in the video.

командир корабля	лётчик
ганторист	бодовой
сворник	танкист

Now you will hear six words or phrases. Check all of those that were in the video.

1.	*4.*
2.	*5.*
3.	*6.*

Translate the following words/phrases into English.

из ОМОНа _____

охотник _____

боксёр _____

Now listen to three words/phrases and translate them into English.

1. _____

2. _____

3. _____

[*Note*. Translation test was not on the same page as the recognition test.]

What was your knowledge of these words before today? Circle the corresponding number.

	Never encountered it	Encountered it	Know its meaning	Use it
боксёр	1	2	3	4
охотник	1	2	3	4
ганторист*	1	2	3	4

[*Note.* Words marked with * are non-words (they were not marked in the original test).]

APPENDIX D. Final Questionnaire

1. Did you like watching Russian videos? Yes ___ No ___

Why?

2. Did you learn any new words when watching the videos? Yes ___ No ___

If yes, describe what you did to figure out the meanings of new words.

3. For each of the following statements, please circle the option that applies to you.
    a. When watching the videos, I was listening to the sound

        All the time Most of the time Half of the time Some of the time Not at all
    b. When watching the videos, I was reading the captions

        All the time Most of the time Half of the time Some of the time Not at all
    c. When watching the videos, I was paying attention to the visual images

        All the time Most of the time Half of the time Some of the time Not at all

4. What were the most difficult things for you when watching the videos?

5. What would have helped you to understand the videos better?

6. What was difficult while you were doing the exercises after the videos?

7. Please rate the following statements on a scale from 5 to 1.
    a. "The captions helped me to understand the videos."

        Strongly agree Strongly disagree

        5 4 3 2 1

    b. "The audio helped me to understand the videos."

        Strongly agree Strongly disagree

        5 4 3 2 1

    c. "The visual images helped me to understand the videos."

        Strongly agree Strongly disagree

        5 4 3 2 1

8. If you have more comments, please write them here.