# Input Variables in Caption Research

**Randall COTTEN**

Abstract

Much research has been carried out to determine the possible benefits of using captions in combination with either moving images and/or spoken dialogue. These types of research can be classified in two basic ways. On the one hand are studies which investigate the effect that either the absence or presence of captions (textual input) have on listening comprehension when subjects are exposed to both spoken dialogue (aural input) and moving pictures (kinetic input). On the other hand are studies which isolate the effect of textual input on listening comprehension in the absence of any kinetic input. In this paper, a review of the different studies will be presented, and suggestions will be given for what other types of variables should be considered.

Keywords: EFL, captions, information processing

## 1.    Introduction

The great popularity of English language movies in Japan is a strong impetus for trying to determine what benefit they might offer to the learner of English as a foreign language (EFL). Japanese learners of English frequently express their enjoyment of watching movies in English, both for gaining exposure to foreign cultures and improving their English. Yet, while EFL teachers may quite readily accept the fact that movies are a plentiful source of material for teaching about foreign cultures and introducing a wide range of metalinguistic features, they might call into question the notion that learners can actually improve their language skills through the casual viewing of films.

Movies might be mistakenly conceived as a passive activity. However, even casual movie viewing is no more of a passive experience than reading is. The viewer/reader interacts both intellectually and emotionally with the scenes on the screen or the words on the page. While many differences could be pointed out about both the similarities and differences between watching movies and reading books, what we wish to focus on here is what type of input is available to the EFL learner when watching English movies. Specifically, there are three main types of input made available in movies: aural (the spoken dialogue); textual (the written dialogue); and kinetic (the moving picture)[1]. One may also argue that, if available, the music which accompanies a scene is another source of input. This will be briefly discussed below.

Given these three main types of input, the question remains as to which one specific type or *combination* of types is the most beneficial for learning a foreign language. It may help to start by discussing the "standard" combination of input types. The discussion here will be limited to Japanese learners watching movies in English. The three basic types of input are: (1) the kinetic input (the picture); (2) the aural input, which is the spoken dialogue in English, or second language (L2); and, (3) the textual input, which are the captions written in the viewer's first language (L1), Japanese. While this combination may be thought of as the "normal" type of input that is available when watching a foreign language film, there in fact are many more different possible combinations. As will be seen below, various manipulations of these input combinations have been studied in an effort to determine which type is the most beneficial to EFL learners.

## 2.    Assessing the Benefits of Closed Captions

In the language teaching field, new technological advances from one area are quickly adopted wholesale or adapted so that they suit the the learning environment. Soon after closed caption technology was developed for the hearing impaired and more television programs were transcribed, language teachers saw potential applications for it with the learners they taught. With the aid of a box which could read the caption signals, teachers were able to show the printed dialogue to their learners. Now teachers could give one more

source of input: an exact written transcription of what was being spoken. While closed captioning was originally developing for the hearing impaired, it can also be a great benefit for hearing students as well.

Accompanying the perceived advantages of closed captioned television programs and films, has come much research on determining the benefits it offers for learners. Kikuchi (1997), in his review of the research since the mid 1970s, found a total of 191 articles published in the United States, and 37 published in Japan that examined the use of closed captions in English education. While the majority of the studies reported that the use of captions was beneficial for the learners, results have at times shown that learners have difficulty processing the information due to the overload of input. Captions seem to sometimes act as an interference for some students. Reese (1984) points out that captions can even impede comprehension because, as the learner "jumps" between the aural and textual channels, some information can be lost.

Of interest here is what many studies do not address. Most of the closed caption research focusses on two basic conditions: those of the benefit of using sound, pictures, and captions (SPC) versus the benefit of using only sound and pictures (SP). In other words, aural, kinetic, and textual input versus only aural and kinetic input. Thus, much of the research on the use of captions focusses only on the presence of L2 captions versus the absence of L2 captions. In most cases, the kinetic input is made available to all the subjects, and the only variable is the availabilty of the captions (conditions 1 or 2 in Table 1). In a few cases, some studies did test for an L1 variable (condition 3). Still rarer in caption studies (Yoshino, et. al., 2000) is the exclusion of the kinetic input.

Table 1: Variables in closed caption studies

| | Textual Input | | Aural Input | | Kinetic Input |
|---|---|---|---|---|---|
| | L1 | L2 | L1 | L2 | |
| **1** | | | | | |
| **2** | *(none)* | | | | |
| **3** | | | | | |
| **4** | | | | | |

While it is standard research practice to isolate the variable which one wishes to test, with the sheer number of caption studies that have already been carried out, one wonders why studies which test the same variables continued to be conducted. It would seem obvious that one additional form of input could have nothing but positive results. While some studies do point out the detrimental effect of input overload caused by textual input through captions, in the end, many of the studies have done nothing but cover already well-worn ground. It is almost as if researchers are trying more to refine their experiments than to test which combinations of input would be the most beneficial.

## 3.    Assessing Textual and Aural Input Combinations

In contrast to the vast amount of studies which have examined the benefits of either the absence or presence of L2 captions, very few have gone beyond this narrow range of variables. In fact, only one study (Lambert, Boehler, and Sidoti, 1981) is known by this author to have gone to the lengths necessary to critically judge the effectiveness of using both L1 and L2 captions with various combinations of L1 and L2 aural input. This study is reviewed in detail below. Unfortunately, no similar research is referenced in the Lambert, et. al. study.

Lambert, Boehler, and Sidoti (1981) studied the effects of using various combinations of L1 and L2 captions and dialogue with 370 fifth and sixth grade elementary learners studying French as a second language. Unlike the studies mentioned in the previous section, their experiment did not involve the use of any input from moving pictures. Excerpts from French language broadcasts from the Canadian Broadcast Corporation were used for the aural input source. For certain test groups, L1 or L2 captions were shown on a television screen (with no moving pictures for any group). The study tested for the effectiveness of nine different combinations or, "conditions" of L1 and L2 textual and aural input (c.f., Table 2). The study also examined the possible influence of the language used on the post-test. Some of the subjects were given a post-test in their L1 (English) or the L2 (French). For our purposes here, this condition is not a relevant factor, so it has not been included. The results of the study showed that subjects who viewed captions written in the L2 while listening to a spoken dialogue in the L1 (and given a post-test in the L2) did almost as well as the group who had been given both textual and aural input in their L1.

Table 2: Textual and Aural Input Combinations (Lambert model)

| | Textual input | | Aural input | |
|---|---|---|---|---|
| | L1 | L2 | L1 | L2 |
| **1** | | | | |
| **2** | *(none)* | | | |
| **3** | | | *(none)* | |
| **4** | | | | |
| **5** | | | | |
| **6** | | | | |
| **7** | *(none)* | | | |
| **8** | | | *(none)* | |

Condition 1 was the same as a closed captioned television broadcast in a foreign language: Subjects read the L2 captions while listening to the spoken dialogue in the L2. Condition 2 used only L2 aural input, with no captions in the L1 nor the L2. In contrast, condition 3 used only L2 captions, with no aural input whatsoever. Condition 4 is the "standard" format available when watching foreign language films: The spoken dialogue is in the foreign language, and the captions are in the L1. In contrast, condition 5 (referred to as "reverse subtitling" by Lambert, et. al.) used L1 aural input, while showing the captions in the L2. Condition 6 is the opposite of condition 1, being the same as a closed captioned television broadcast, but in the L1. Condition 7 is the opposite of condition 2, using only L1 aural input. Condition 8, being the opposite of condition 3, used no aural input with only captions in the L1. The one additional condition which has been omitted here was the same as condition 5 except that the post-test was given in the L2. Finally, it should be acknowledged that the purpose for including conditions 6 through 8 in the experiment might be questioned because these conditions only test for comprehension in the L1. However, one must remember that the subjects in this study were fifth and sixth grade elementary students, so it was relevant to examine the overall comprehensibility of the source material in the subjects' native language.

Out of the nine different conditions for L1 and L2 dialogue/script combinations, the most favorable were those in which the L1 was made available through either the written script or the spoken dialogue. One condition, "Reversed Subtitling-L2" (condition 5) revealed unexpected results. When learners listened to the dialogue in their native language (English), and read the L2 (French) captions, they performed almost as well as in condition 1 – listening and reading in their native language. The researchers hypothesized that this was because the learners were able to grasp the overall message in their L1 with little effort, and were then able to see in the captions how to formulate the same expressions in the L2.

While reverse subtitling might at first seem contrary to set notions of using foreign films to learn languages, when we consider that L1 aural input can be processed easily and quickly, then the expressions in the target language can be read in the L2, it does appear to be a very beneficial input combination. "Standard" (L1) captions, in a sense, simply get in the way. Textual processing in the L1 tends to inhibit processing of the more difficult L2 auditory message. Lambert, Boehler, and Sidoti (1981) suggest that L1 aural input enables the learner to use more efficient top-down processing of the L2 textual input. They go on to further suggest that learners who use television programs and movies would benefit much more greatly simply by being able to see the L2 script instead of having it translated into their L1.

Interestingly, this study contradicts findings that L2 captions can create an overload of information, thus impeding information processing and, ultimately, comprehension. Their research results show that the two most effective textual/aural input combinations are conditions 1 and 5. The authors go on to suggest that condition 5 (reversed subtitles) would be a good starting point for less advanced learners and, after they gain more proficiency, should graduate to L2 captions and audio (condition 1).

## 4.    Other Variables for Consideration

The most obvious difference between the study discussed above with the closed caption studies discussed in section two of this paper is the absence of any kinetic input. Lambert, et. al. (1981) begin and end their paper with statements to the effect that they were interested in how textual input could benefit language learners when watching television or films. Considering the breadth and depth of their study, if they had included more test groups which included the presence of input from moving pictures, their number of conditions would have grown so large as to be practically impossible to carry out the experiment and accurately analyze the results. On the other hand, we have seen that the closed caption studies which include a kinetic input element do not

satisfactorily test for the effects brought about by variations in L1 and L2 aural and textual input. As educators, teachers come to this field of inquiry because of an interest in the potential learning benefits that television and film media offer to the learner. As researchers, it is important to isolate and test the relevant pieces of the puzzle that are involved in this process. Yet, before this can be achieved, it is important to consider what kind of variables have not yet been addressed fully in this area of research.

Research which sets out to study the effects of textual input needs to consider the effect of the speed at which the subjects are required to read, and the extent to which the kinetic (and, possibly even musical) input has on comprehension ability. Some studies (e.g., Yoshino, et. al., 2000; Markam, 1989) do mention the number of words per minute which the spoken dialogue (thus, also the captions) contained. A possible factor to consider is whether subjects are able to listen and comprehend at the same rate as they read. While calculating the number of words per minute in a given section is important, the rate at which a person speaks will vary widely according to the type of media. For news broadcasts, or prepared materials, we might expect the rate of speech to remain fairly consistent. However, when using films or television programs, it should be remembered that some scenes will have faster rates of speech, others slower. If these scenes are shown with captions, the text will remain on the screen for shorter or longer durations of time. All of these factors are going to affect the comprehensibility of any given passage. It is not a simple matter to say whether captions alone are going help or not. When looking at the results of such research, one needs to know not only how many words on average are spoken and presented in text, but also whether or not the presentation speed remained consistent throughout. This is closely related to how fast a learner is able to read and comprehend the captions, which are usually presented at a speed equivalent to the rate of speech. Studies concerning the degree to which varying time exposures to textual input has on comprehension ability will be informative to understand the connection between caption presentation speed and reading speed and, ultimately, the benefits of using captions in combination with kinetic and aural input.

On another point related to media types, TV programs and films make great use of music to create mood, build tension, provide clues to what might happen next, or give insights into the feelings or intentions of the characters. Yet,

while background music may be categorized under aural input, it may or may not be directly related to the spoken input. Whimsical music might suggest that, *contrary* to what is being said, the character is lying or playing a trick on someone. Furthermore, as in real life, characters in movies do not always finish what they are saying. The music that fades in clues the viewer into what might be said next. Foreboding or sad music can fill in information that is otherwise left to our imagination. The point here is that we should not ignore the benefit that background music may provide. A movie which is absent of music would be very flat, and seem to lack emotion or be too tedious due to the need to "spell out" everything. Thus, another important consideration in caption comprehension studies which involve the use of excerpts from movies or TV programs is to what degree background music may provide information for deriving meaning from the context in a scene.

Almost consistently, studies do not mention anything about the difficulty of vocabulary and sentence structures in the materials used. Researchers explain the general level of the subjects, but write nothing about whether they could actually be expected to know and understand the words that they would encounter. At the heart of this issue is the act of reading, yet the reading of captions must be separated from the "normal" reading of words and sentences printed on a page. Seeing captions is perhaps closer to the act of listening than it is that of reading. The words are shown once, then gone again. The viewer has no chance to go back and refer to an earlier part in the text. This type of test design requires subjects to be able to almost instantaneously read and process the information seen in the captions, and then derive meaning from it. Not only do subjects need to rely on their ability to take in printed material, but they also need to do it very quickly. Testing the affects of textual input through captions is also testing the ability to speed read. We might need to ask to what extent much research might actually be only testing the subjects' ability to read and understand simultaneously. It would appear that the very act of reading captions (i.e., degree of ability) should be treated as a separate research concern in itself.

Another issue which needs to be considered is the relation of the content of the spoken message to the visual context. When testing the affect of seeing only the picture versus seeing the picture with captions, the scene itself needs to be judged according to how well the words match what is

being shown in the scene. Again, this is where the type of media is highly influential on the results. The spoken dialogue for in documentaries, for example, would most likely be very closely related to what is being seen (the kinetic input). However, in movies and television programs, there is a likelihood that the spoken dialogue in some scenes could have very little relation to what we are looking at. For example, if two speakers are sitting in a coffee shop talking about something that happened in the past, the only important kinetic information would be the reaction on the speakers' faces showing their emotion in relation to the dialogue. While a person's facial reaction is closely related to the content of the spoken words and important information for understanding what was said, we must acknowledge that the dialogue in such scenes has relatively little support from the visual information. On the other hand, in a scene in which a father is teaching his young son how to hit a baseball, the words and the action will closely match one another. In ranking this aural-kinetic relation, we might borrow the term which Reese (1984) used, "visual-verbal redundancy". Thus, the former scene could be said to have low visual-verbal redundancy quotient, while the latter scene could be said to have a high quotient of visual-verbal redundancy.

The features discussed here are all closely related, and any one of them may have an effect on the overall results of a study which intends to examine the effects of captions when used in combination with audio or motion picture input. The features which should be considered when selecting the most appropriate type of media have been put into a matrix form in Table 3, below.

Table 3: Variability Factors in Media Types

| Media type | Content |
|---|---|
| • authentic vs. prepared<br>• movie or TV program<br>• news broadcast<br>• musical support | • words per minute (quantity)<br>• level of difficulty |
| Caption/Dialogue Speed | Visual-verbal redundancy |
| • consistently slow or fast?<br>• some parts slower or faster than others? | • high quotient: pictures support dialogue<br>• low quotient: pictures do not match dialogue |

It is useful to see these elements as belonging to matrix because each component may be closely related to another

one. The above matrix can help inform researchers' decisions. For example, while it may seem best to use an authentic TV news broadcast in which announcers use a fairly consistent rate of speech, their speech is often quite fast, and pictures usually often do not have a high visual-verbal redundancy quotient. Furthermore, the level of difficulty may be too high for all those but the most advanced learners. Using such a matrix, source material can be carefully judged and selected according to how these factors interact with each other. Accordingly, during the analysis of the results, these features should also be taken into consideration.

## 5.  Conclusion

The research that we have been looking at here aims to find the benefits of using printed captions in combination with other types of input (sound and/or moving images). The stated benefits are how the captions can aid comprehension, which can be either an overall understanding of the content of the material, or better listening comprehension. Many critisicms have been raised here about these studies. However, before embarking on a research project, a review of the available literature is usually the first step. In this area of enquiry, more questions have arisen than have been answered, and most of them are related to what was not tested.

This type of research involves a wide variety of issues: listening comprehension, speed reading, information processing, and multiple input types. Each of these areas alone is an important part of the learning process. Caption research, in a sense, tries to discover how they interact (either positively or negatively) with each other. The fact remains that there are many complicated parts of the puzzle of how to best test the benefits of using captions in combination with input from spoken dialogues and moving pictures. Careful isolation and critical evaluation of each component is essential in order to arrive at results that will be beneficial to learners and teachers across a wide range of learning environments. It is hoped that the questions raised and suggestions offered in this paper will be of use to future research.

## Notes

1.  The term *kinetic* is used rather than *visual* in order to avoid confusion with captions, which could also be thought of as "visual input". (An alternative to kinetic could be the term "pictorial".) In any case, *kinetic* will be

used here to distinguish between what is *seen* in the movie from what is *read* in the captions (textual input).

## References

Kikuchi, T. (1997). A review of research on the educational use of English captioned materials in Japan. *Caption Central*. Available: http://captioning.robson.org/resources /papers/kikuchi. html [23 Sept. 2003].

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.

Markham, P. L. (1989). The effects of captioned television videotapes on the listening comprehension of beginning, intermediate, and advanced ESL students. *Educational Technology*, 29 (10): 38-41.

Reese, S. D. (1984). Visual-verbal redundancy effects on television news learning. *Journal of Broadcasting*, 28 (1): 79-87.

Yoshino, S., Kano, N. & Akahori, K. (2000). The effects of English and Japanese captions on the listening comprehension of Japanese EFL students. *Language Laboratory*, 37: 111-130.

16    11    26