

A Case Study of Visual-verbal Relations and Application Principles in China's College English Classroom

Peipei Yang¹

¹ Foreign Languages School, Zhejiang Gongshang University, Hangzhou, China

Correspondence: Peipei Yang, Foreign Languages School, Zhejiang Gongshang University, Hangzhou, 310018, Zhejiang, China. E-mail: ypppink@hotmail.com

Received: August 1, 2014 Accepted: October 19, 2014 Online Published: November 25, 2014

doi:10.5539/ijel.v4n6p104 URL: <http://dx.doi.org/10.5539/ijel.v4n6p104>

Abstract

This study seeks to explore relations of visual-verbal modes and figure out application principles in China's College English Classroom (CEC). It takes data from two files: (1) videos of two excellent CEC teachers; and (2) semi-structured interviews with them, within which it studies four modes in PPT or on blackboard presentation: image, words, dynamic and symbol. Three instruments—Multimodality annotation software ELAN, two-dimensional meaning-making tables and semi-structured interview are employed to facilitate both quantitative and qualitative analyses. The results showed features of frequency, timing and proportion of each mode summarized by ELAN and found out proper collocation of image, words, dynamic and symbol relies on intersemiotic relations, which are revealed as complementary and non-complementary. It further analyzed application principles of four modes under CEC context in China. Except principles of effectiveness, efficiency and appropriate collocation put forward by former studies, it complemented principle of modes' transference to highlight the necessity to form students' autonomy of visual-verbal modes.

Keywords: visual-verbal modes, College English Classroom, application principles, multimodality

1. Introduction

1.1 Context of the study

Visual-verbal studies of TESOL (Teaching English to Speakers of Other Languages) originate from Western countries and are springing up among Chinese researchers nowadays. In conformity to the growing trend of applying multimodal pedagogy in language teaching, in 2004, the Ministry of Education of the People's Republic of China published the trial version of *College English Curriculum Requirements* for non-English majors as well as for us teachers across colleges and universities. As the requirements stated, due to the rapid growth of the number of university students and a relatively limited quantity of educational resources, the Ministry calls on universities to make full use of the opportunities that multimedia and Internet technology have brought to us. At the same time, a new pedagogy based on modern information technology should substitute for traditional mono teaching model, in the hope of highlighting students' dominant status in teaching processes.

Dated back to Halliday's social semiotic theories, the following multimodal studies have been carried out mainly based on modes' meaning-making from a systemic-functional perspective. The New London Group (2000), firstly put forward "multiliteracies" that aim to equip students with literacy for various modes, including visual-verbal ones. Kress and van Leeuwen (2006) focused on visual modes and created a metalanguage for them, that is, "Visual Grammar". Royce (2002, 2007) also mainly examined visual modes in TESOL and explored visual-verbal synergy in the classroom. Other researchers such as Baldry and Thibault (2006), Bezemer (2008) and Cloonan (2011) not only studied visual modes but also spatial modes, gestural modes or audio modes in the TESOL classroom.

Compared with Western studies, Chinese scholars began to have some achievements in this field in recent years. For example, Hu and Dong (2006), Chen and Qian (2011) and Zhang (2010, 2012) all have contributed to visual-verbal studies in TESOL. Under the general purpose of achieving optimal result, three principles for the choice of modality are proposed: effectiveness principle, efficiency principle and principle of appropriate collocation (Zhang, 2010). Effectiveness principle suggests that the choice of any mode takes better teaching effectiveness as a prerequisite so as to avoid the mode useless or whose negative effect overweighs positive one

in certain conditions. Efficiency principle originates from the conflict between optimization and simplification, thus requiring a balance between them to be bore in mind while choosing among multiple modes. When it comes to the distribution of multiple modes, principle of appropriate collocation considers mutual cooperation among different modes under the standards of optimal match.

Based on the demonstration of research background, though visual-verbal elements are relatively controversial and heatedly discussed by researchers, they are usually discussed under general TESOL circumstances. While in fact, teachers teaching College English to non-English majors form the largest group among all the TESOL teachers in China. Therefore, it is key to specify visual-verbal studies in the compulsory course of College English for non-English majors in China.

Under the demand of “hypertextual” pedagogical concept from China National “*College English Curriculum Requirements*” (Wei, 2009: 140) and through our pilot studies about CEC (College English Classroom), visual-verbal modes in CEC mainly consist of images, words, dynamic and symbol in teachers’ PPT or their blackboard layout. What is more, in the empirical process of CEC, visual-verbal modes are relatively convenient and appropriate to observe or mark (Schmid, 2008) compared with gestural modes or spatial modes which are relatively more influenced by personal reference. In another word, visual-verbal ones tend to be more organized with rules that we can abide by so that visual-verbal meaning-making and relative applied principles are persuasive for applying other modes to CEC.

From the point of demonstration effect and research feasibility, two excellent CEC teachers A and B from a provincial key public university are selected as two cases. To avoid subjectivity, they are determined by teaching awards or honorary titles they have won and their experiences or duties in their department. A is the first-prize winner of the provincial final round of the 1st SFLEP National Foreign Language Teaching Contest, of which SFLEP is short for Shanghai Foreign Language Education Press (official website: <http://nfltc.sflep.com/>). In addition, A is the director of CEC Teaching and Research office in that university and has more than ten years’ experience of CEC teaching. At the same time, B is the second-prize winner of the provincial final round of the 2nd SFLEP National Foreign Language Teaching Contest. Her annual evaluation performance by this university is outstanding among all the teachers for students of her CEC class speak highly of her teaching, thus having been invited many times as lecturer for CEC teaching seminar. Correspondently we selected A’s contest clip from SFLEP National Foreign Language Teaching Contest and a video of B’s on-spot class from our recordings, of which the details will be elaborated on in the next chapter. Later video A and video B have been analyzed and semi-structured interview to teacher A and teacher B have been conducted. Multimodality annotation software ELAN is adopted to demonstrate the features of frequency, timing and proportion of the visual-verbal modes in two cases; two-dimensional meaning-making tables are designed to discover some characteristics about the distribution and intersemiotic relations of images, words, dynamic and symbol in PPT or on blackboard and semi-structured interview are employed to facilitate both quantitative and qualitative analyses.

In order to help CEC teachers to better design, select and distribute visual-verbal modes, his study further intends to justify whether Zhang’s three principles mentioned before for choosing modes work under CEC context and explore any new principles applicable to CEC if possible.

1.2 Purpose of the Study

Based on the context of the study, three main research questions are identified as follows. These questions focus on four visual-verbal modes in two cases we select: image, words, dynamic and symbol. The first question is quantitative-oriented, and the rest two are qualitative-oriented:

1. As to the four major visual-verbal modes studied in cases — image, words, dynamic and symbol, what are the features of their frequency, timing, and proportion?
2. How do the two excellent teachers collocate these four major visual-verbal modes in meaning-making of CEC teaching and what are the intersemiotic relations of these modes?
3. Are Zhang’s principles for the choice of multiple modes — effective principle, efficiency principle and principle of appropriate collocation — specifically appropriate for visual-verbal modes under CEC context and is there any complementary principle we can abide by based on the results from the first two questions?

Answers to the above three questions will be further discussed to reveal the enlightenment that this study of visual-verbal meaning-making and application principles brings to CEC teaching or other relevant studies in this field, especially to the real processes of designing, selecting and distributing visual-verbal modes.

2. Methods

2.1 Data Collection

As there are two types of data collected for this study—video data and transcription data, this section shows the processes of data collection respectively.

As to video data, before we settle down to the two videos A and B, we have done some pilot studies mainly about video selection and recording. In the beginning period, we watched many contestants' clips from SFLEP series and repetitively tried to annotate some of them with ELAN, by which the skills of observing CEC teaching and annotating videos have been gradually improved. At the same time, we attended CEC of several teachers in a provincial key public university and recorded their classes. Finally Video A is selected as A's CEC teaching excerpted from the national final round of the 1st SFLEP National Foreign Language Teaching Contest. The classroom content is concerned with Unit 6, Book Three of *New College English* published by Zhejiang University Press and this whole class lasts about 19 minutes. Video B is one of the on-spot records of B's CEC teaching by non-participant observation during pilot study. The classroom content is concerned with Unit 3, Book Four of *New College English* published by Zhejiang University Press and this whole class lasts about 30 minutes. The titles of two units are *Man and Animals* and *Gender Difference* respectively.

As complementary data, transcription data are collected after analyses of video data since the semi-structured questionnaire is designed based on the results from video analyses. Then A and B are invited to participate in the semi-structured interview separately and two interviews are both recorded for further transcription. Considering language-friendliness, both the questionnaire and the interviews are performed in Chinese.

2.2 Discussion

Validity and Reliability of cases are responsible for guarding against criticism or flaws in case study. Therefore, it is necessary to analyze validity and reliability in this study. Validity means that “identified codes can explain the data” or “the codes should be fully supported by the data rather than something that the researcher has imposed on the data” (Wen, 2004: 248). Firstly, since A and B are excellent in applying visual-verbal modes to CEC as analyzed in section 1.1 and both of their videos involve observable visual-verbal modes, visual-verbal relations reflected by these codes are relatively demonstrative and can support the discussion of visual-verbal application principles. Secondly, quantitative data from ELAN that elaborate on the modes' annotation and statistics of frequency, timing, and proportion to some extent help triangulate the validity of the data collected for this case study. As to reliability, there are several means that can ensure it. The major three ways to confirm reliability include: “repetitive reading of and continuous thinking about the data” by the researcher; two or more researchers' cooperative work to “obtain similar codes” while “working at the same set of data”; or invitation to “a third party” to co-participate in the analysis (Wen, 2004: 248). In accordance with the first approach, through a series of pilot study, such as other contestants' clips analysis from SFLEP, repetitive ELAN annotation and counting on trial to video A and B, and many times of non-participant observation of B's on-spot CEC, we try to avoid occasionality but improve accuracy. Besides, A and B's participation through semi-structured interview goes in line with the third way, serving as a complementary research tool for providing communication between the researcher and the third party.

2.3 Instruments

Three instruments—multimodal analysis software ELAN, two-dimensional tables of visual-verbal meaning-making, and semi-structured interview to teachers of our cases—are adopted, of which the first two are dedicated to video data analysis and the third one is for collecting transcription as complementary data. To elaborate, ELAN is designed to process sample videos of a complete class while two-dimensional tables are intended to analyze certain typical visual-verbal teaching links. Based on the results from ELAN and sheets, the semi-structured interview questions are designed to obtain A's and B's feedback as complementary data. As to the definition of “image”, “words”, “dynamic” and “symbol” in PPT or on blackboard, it should be clarified here as an annotation standard (see Table 1).

Table 1. Annotating “image”, “words”, “dynamic” and “symbol”

Visual-verbal modes	Instruction for annotation
Image	A visual presentation of an object or scene, an animal or a person, or some abstraction
Words	All the words no matter separately from or synergistically with visual modes
Dynamic	Kinetic performance or characteristics of images if they are not static
Symbols	Visible statistics, arrows, tables and frames

ELAN (EUDICO Linguistic Annotator), which can be downloaded from the ELAN official website (<http://tla.mpi.nl/tools/tla-tools/elan/>) equipped with user guide, is one of the most practical and popular multimodal analysis software. It is a professional annotation tool that “allows you to create, edit, visualize and search annotations for video and audio data”, and “specifically designed for the analysis of language, sign language, and gesture” (User Guide for ELAN Linguistic Annotator, 2012: 2). The version we use is the recently updated one in 2013 or ELAN v 4.5.1.

In this study, four steps can be summarized as the annotation and counting processes: defining linguistic types and tiers, selecting time intervals in each tier, entering annotations and at last ELAN automatic counting. While creating tiers, every tier needs to be named; therefore four tiers are named after their abbreviations as “im”, “wo”, “dy” and “sy”. Figure 1 and Table 2 and 3 respectively show the interface of annotation and statistics.

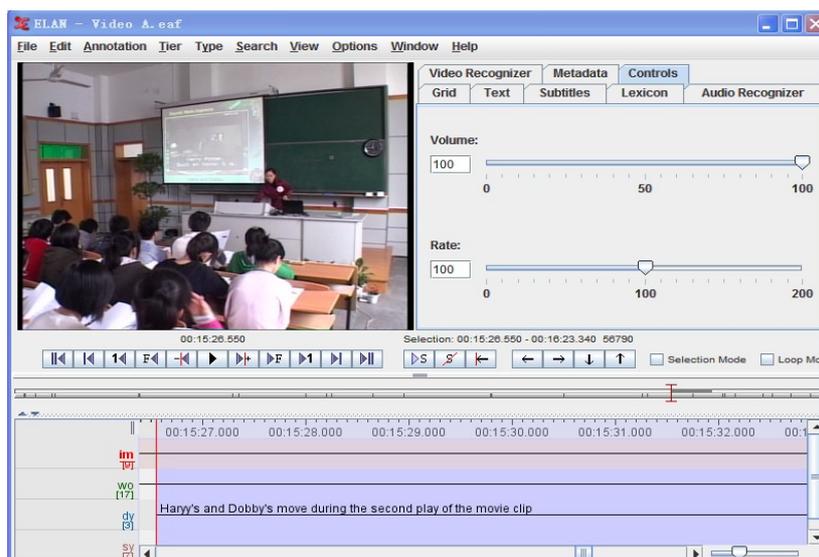


Figure 1. Time interval selection and annotation entering

Note. Video A is taken as an example in this display window.

Table 2. ELAN annotation statistics of video A

Tier	Number of annotation	Minimal duration	Maximal duration	Average duration	Median duration	Total annotation duration	Annotation percentage
im	9	7.05	183.04	62.085556	28.97	558.77	48.921
wo	17	7.06	189.53	57.864292	38.636	983.693	86.124
dy	3	8.88	56.79	39.00333	51.34	117.01	10.244
sy	7	33.54	204.63	81.812286	40.53	572.686	50.14

Table 3. ELAN annotation statistics of video B

Tier	Number of annotation	Minimal duration	Maximal duration	Average duration	Median duration	Total annotation duration	Annotation percentage
im	5	8.9	169.31	90.428604	65.813	452.143	24.639
wo	19	1.11	259.73	79.702423	72.866	1514.346	82.523
dy	4	8.91	161.17	90.262497	95.485	361.05	19.675
sy	4	100.5	259.72	150.119995	120.13	600.48	32.723

Complementary to ELAN annotation of the two whole videos, in order to analyze some typical visual-verbal teaching links from video A and B, a matrix with three types of meanings (Kress & van Leeuwen, 2006) on the vertical axis (representational, interactive, and compositional) and the four modes in this study on horizontal axis (image, words, dynamic, and symbol) are used. From the fulfillment of these tables, we can summarize common

visual-verbal relations reflected by these samples and contribute to the discussion of application principles from the perspective of teaching links. Table 4 processes the lead-in part of video A as a sample and other 3 samples of different teaching links are listed and analyzed in this study.

Table 4. Sample 1: two-dimensional table of visual-verbal meaning-making

Dimension of meanings	of	Dimension of visual-verbal modes (image, word, dynamic, symbol)
representational		<i>im</i> : Real object images of fish and bear generate ideology of fish and bear.
		<i>wo</i> : “Unit 6 Man and Animals” written in art font represents the title of the unite.
		<i>im+dy</i> : Images with dynamic effects strengthen ideological meaning of the two kinds of animals, making them more vivid.
Interactive		<i>im+dy+wo</i> : Dynamic images and words turn up together: words illustrate dynamic images and images correspond to representational meaning of words.
		<i>wo</i> : In the words “Man and Animals”, “and” constructs relationship between students and animals.
		<i>dy+im</i> : The jumping of fish and hunting by bear arouse students to have some feelings or attitudes, thus interacting with students.
Compositional		<i>im</i> : Fish and bear are in diagonal angle of view: big bear occupies top-left; tiny fish occupies bottom-right.
		<i>dy</i> : Dynamics highlight images as the main content of a whole PPT page with all the rest static.
		<i>wo</i> : The title is located in the top-middle with an inclusively leading effect.
		<i>im+dy+wo</i> : Dynamic fish and bear occupy the center and most space with static background; the title is located in the top-middle.

Respectively based on analyses of video A and B, questionnaire A is dedicated to teacher A and questionnaire B to teacher B. Every questionnaire includes 6 to 7 open questions about typical visual-verbal modes captured from videos, without any choices presented. The questions are used to elicit the visual-verbal modes’ usage intention of A and B and their opinions about how to carry out these processes under real CEC context. Neither is the theme of this study exposed to A and B, nor do any key terminologies be involved during the semi-structured interview so as to guarantee the spontaneity of their feedback.

2.4 Data Analysis

From statistics in “Number of annotation” as shown in Table 2, “Im”, “wo”, “dy”, “sy” are respectively annotated 9, 17, 3 and 7 times; in Table 3, they are respectively 5, 19, 4 and 4 times. It can be found out that words appear most frequently, followed by image and symbol with dynamic relatively turning up least, which suggest that we should apply these four modes effectively to where they are needed under the diverse and dynamic CEC context.

From statistics in “Total annotation duration” and “Annotation percentage” verbal mode—words in this study—still occupies the largest one, and next to it are similar proportions of image and symbol, and in the bottom is dynamic with the least proportion. From these timing features, we get a hint that these four modes should be used in consideration of economical effects, or in other words, efficiency.

From statistics in the above two tables, continuous observation of a variety of time intervals in timeline viewer and annotation text in specific viewer, we conclude that at least more than half of words appear in synergy with image, dynamic or symbol. It can also be summarized that although all the dynamics appear in synergy with image or symbol, there are still some static images or symbols. Synergy does not mean that modes have to begin and end at the same time; instead, they usually appear successively with certain overlap. The implication drawn here is that appropriate collocation of these four modes should be kept in mind as a principle.

Under analysis of two-dimensional tables, except Sample 1 in Table 4, Sample 2 is a dubbing activity (see Table 5) for Harry Potter and a dog named Dobby. On the other hand, since video B is about Gender Difference, hence Sample 3 is the discussion practice of gender difference between man and woman (see Table 6) and Sample 4 is an extended listening exercise (see Table 7) about gender equality.

Table 5. Sample 2: two-dimensional table of visual-verbal meaning-making

Dimension of meanings	of	Dimension of visual-verbal modes (image, word, dynamic, symbol)
representational		<p><i>wo</i>: Film subtitles represent content of the conversation in the movie clip while the name of the speaker represents who is speaking right now.</p> <p><i>im</i>: Images of Harry and Dobby represent two main characters of the clip, here specifically symbolizing man and animals respectively.</p> <p><i>im+dy</i>: Dynamic images of Harry and Dobby represent their gestures, actions or facial expressions, etc.</p> <p><i>im+wo+dy</i>: All the film subtitles work in concert with dynamic images to represent the plots as well as characters of this clip.</p>
Interactive		<p><i>wo</i>: What the students think about when they watch or read subtitles, for example, they are most likely to sense Harry's politeness and friendliness towards Dobby.</p> <p><i>im+dy</i>: Here refers to what students feel about Harry's and Dobby's gestures, actions or facial expressions, etc. For example, they probably feel that Dobby appreciates Harry's attitude towards him very much.</p>
Compositional		<p><i>wo</i>: Utterance subtitles are showed in the bottom of the screen for watching and dubbing while the name of the current speaker is located in the top-middle to indicate whose turn it is for dubbing.</p> <p><i>im+dy</i>: Dynamic makes images alive through forms of gestures, actions or facial expressions, etc.</p> <p><i>im+wo+dy</i>: Dynamic images occupy all the background, giving an impression that visual modes are dominant here compared with verbal words as subtitles.</p>

Table 6. Sample 3: two-dimensional table of visual-verbal meaning-making

Dimension of meanings	of	Dimension of visual-verbal modes (image, word, dynamic, symbol)
Representational		<p><i>im</i>: Head portraits of man and woman generate concepts of two kinds of human beings.</p> <p><i>wo</i>: Independent chunks of words description generate representative human beings' characteristics, whether it is feminine or masculine is unknown yet.</p> <p><i>sy</i>: Statistics symbols mean numbers from investigation practice and arrows mean the concept of match or relation.</p> <p><i>im+wo+sy</i>: Word chunks, head portraits and arrows connecting them present which feature belongs to which gender. For example, "pay \$ 200 for a \$ 300 item needed" is typical for males while "pay \$ 300 for a \$ 200 item not needed" fits females.</p>
Interactive		<p><i>wo</i>: Word descriptions stimulate students to relate them to real life.</p> <p><i>im</i>: Man's and woman's portraits interact with students' inner concept about gender.</p> <p><i>sy+wo+im</i>: They cooperatively comprise the matching exercises to students.</p>
Compositional		<p><i>im</i>: Head portraits are displayed in the center.</p> <p><i>wo</i>: Word chunks are randomly distributed around the center.</p> <p><i>sy</i>: Arrows are between chunks and portraits and statistics are within chunks.</p>

Table 7. Sample 4: two-dimensional table of visual-verbal meaning-making

Dimension of meanings	of	Dimension of visual-verbal modes (image, word, dynamic, symbol)
Representational		<p><i>wo</i>: Question list in the second time of listening and blank filling in the third time represent meanings of questions.</p> <p><i>im+dy</i>: Dynamic images in the news video offer representational meaning in the first time of listening.</p> <p><i>im+wo+dy</i>: Switches between dynamic images and verbal words or within verbal words strengthen students' understanding of the listening material.</p>
Interactive		<p><i>wo</i>: Questions interact with students through the form of inquiry.</p> <p><i>im+dy</i>: Here refers to what students feel about dynamic images shown in the news.</p>
Compositional		<p><i>wo</i>: Subjective questions are listed sequentially on the entire PPT page according to the news development in the second time of listening and blank fillings is demonstrated in a word document on another PPT page in the third time.</p> <p><i>im+dy</i>: The video is played in full screen without any subtitles for the first time while it is invisible for the second and third times.</p>

Note. In this teaching link, B plays three times of the extended listening practice, with the first time with the video screen on the projection, the second time with a question list and the third time with a blank filling document.

From the above table analyses of four typical visual-verbal teaching links, we can summarize some aspects of visual-verbal relations through their meaning-making in CEC context. Firstly, words and image are the dominant modes in generating representational meaning; it is common to see they are accompanied by dynamic and symbol, which may act as assistance methods to strengthen and complement representational meaning or just play compositional roles in the background. Secondly, symbols such as frames, arrows, take clearer visual effects by presenting information in a more organized and intuitive way. Thirdly, in terms of interaction, dynamics work not to strengthen or complement meaning but to exist together with images or words to interact with students. Finally, switches between visual-verbal modes with hypertextual functions can benefit students' understanding and multiliteracies cultivation, such as listening practice in video B and dubbing exercises in video A.

The examination of transcriptions tries to find out whether feedback from A and B matches Zhang's principles for the choice of modes or even gives some hint for any complementary principle.

Firstly, both A's and B's answers indicate that almost all the visual-verbal modes applied originate from certain teaching goals, of which the effectiveness to realize these goals is given priority to when they are considering and comparing the designs during their class preparation. As teacher A responded in our interview, some words presented with pictures of man and animal can relieve or prevent visual fatigue for students, while overuse of pictures leading to visual fatigue and even distraction are not advocated. Therefore, some visual-elements can be composed in the background. In this aspect, B also expresses that she selected visual-verbal modes first and foremost in consideration of the specific CEC context. For example, for the second and third time of listening, B hides video to avoid distraction to students and verbal blank fillings offer students a framework to depend on to capture details of the listening material, especially some phonetic difficulties as liaison and out of explosion.

Secondly for the sake of time planning both during their pre-class preparation and actual proceedings of class, they all consider about economical efficiency, which is consistent with Zhang's efficiency principle. A suggested we control video time in CEC within 2 or 3 minutes as the dubbing video in case A is about 1.5 minutes. She summarized her reasons as two points: firstly since time of CEC is limited, we need to develop all the means to activate more students' participation instead of spending too much time in video-watching; secondly time limit of video can help balance the difficulty level. On the other side, B put 3 to 5 minutes as the time limit of videos in CEC, emphasizing that too long audio-video materials are not good for students' understanding and memorizing.

Thirdly, as to collocation of visual-verbal modes, the fact that A and B mentioned quite a few times about this point goes in line with Zhang's principle of appropriate collocation. A emphasized that mere PPT text poses more pressure on students especially when too many words are typed on one PPT page. Proper combination of texts with tables, arrows, images or statistics not only relieve the pressure but also leave more space for teachers to play. As B stated, she pays great attention to cohesion and coherence during modes' collocation. For instance, full screen videos accelerate students' adaptation to the listening material when they first approach it.

Finally from communicating with them, we notice that they both highlight students-centered teaching style and value students' autonomy of modes very much, which is not able to belong to any of Zhang's three principles and has provided much inspiration for seeking a complementary principle. Moreover, this point coincides with modes' transference presented from table analysis. As A explained, the dubbing teaching link is distributed before the conclusion of the class like a prelude to sublimation part of the entire class, which offers a good opportunity for students to practice oral English and fully comprehend relations of man and animals. B explained that activities such as matching exercises aim to leave time and space for students' thinking instead of accepting information passively. She jokingly said that surprises will be given if there are some gaps between keys and students' own answers and if they are similar, students will feel sympathetic.

3. Results

Results of ELAN, tables, and semi-structured interview combine the researcher's findings with opinions of A and B as research subject and practitioner, revealing visual-verbal relations and application principles of designing, selecting, and distributing modes, are to some extent enlightening for CEC teaching.

On the basis of ELAN analysis of effective application of four modes in consideration of time efficiency and collocation appropriateness, two-dimensional tables of visual-verbal meaning-making have identified more details about visual-verbal relations, which can be mainly divided into two categories: complementary and non-complementary, obviously becoming evident for the principle of appropriate collocation.

Complementary relations refer to one mode's reinforced or complementary effect to the dominant and basic mode under that context. That is to say, the dominant mode can work without other modes' accompaniment but

the effect will be strengthened with the contribution of the complementary relations. For example, as the dominant modes in generating representational meaning, dynamic and symbol accompany words and image to strengthen and complement representational meaning or just play compositional roles in the background. Non-complementary relations refer to the co-existence within which one mode plays similar or opposite roles to cooperate with another, instead of helping strengthen the same meanings. To put it more vivid, some of modes' non-complementary relations can be compared to "analogy" or "antiphrasis" in linguistic terms.

Through data analysis it is found out that though visual-verbal modes follow Zhang's tree principles, there is still one tip to be highlighted. That is, transform from teacher-orientedness to students-centeredness by transferring autonomy of modes to students. Principle of modes' transference means that visual-verbal modes originally designed by teachers should be autonomized by students, such as the dubbing activity and the word-picture matching exercises in this study, thus benefiting students' understanding and multiliteracies cultivation.

4. Findings and Limitations

To conclude, based on three theoretic frameworks as social semiotic theory, visual grammar, and Zhang's principle approach, the present study takes two excellent teachers as participants and their CEC videos and interview transcriptions as data, and explores visual-verbal relations during meaning-making processes and their application principles under dynamic CEC context. The three major research findings are sequenced here to mirror the order of the three research questions in 1.2.

Firstly, as to image, words, dynamic and symbol on the PPT of cases, synergy does not mean total simultaneity but means a certain overlap. Statistics from ELAN, which show features of frequency, timing, and proportion of the visual-verbal modes in CEC, seem to provide some rules: "words" and "image" rank first and second in proportion higher than "symbol" and "dynamic" on PPT or blackboard, with usually more than half of "words" in synergy with visual elements and some "image" and "symbol" in synergy with "dynamic". These rules are evident for the consideration of effectiveness and efficiency while applying visual-verbal modes in CEC and offer references for distributing modes and arranging the time they occupy.

Secondly, results from two-dimensional meaning-making tables and semi-structured interviews prove that the appropriate collocation of image, words, dynamic and symbol, which relies on intersemiotic relations privileges two cases and is highly valued by these two excellent CEC teachers. The intersemiotic relations of the four major visual-verbal modes in meaning-making of CEC teaching are revealed as either complementary or non-complementary. Hence the distribution of visual-verbal modes on PPT or blackboard should take these intersemiotic relations into consideration.

Thirdly, the analysis of cases as evidences used to testify Zhang's three principles of multiple modes in foreign language teaching—effective principle, efficiency principle and principle of appropriate collocation—are also applicable to visual-verbal modes under dynamic CEC context. Besides, there is one complementary application principle yielded by analyses that we did not plan to testify originally but later proved to be evident through the whole study, that is, principle of modes' transference which acts as a reminder to transfer teacher-dominated modes' application to students' autonomy of modes.

The implications and enlightenments of this study are summarized here from three aspects: theory, methodology, and practice. Theoretically, it lies in the complementary principle -principle of modes' transference-to Zhang's principle theory. The methodological implication mainly refers to the possibility that some of the research instruments here might also apply to other multimodal studies. For example, ELAN, as professional annotation software dealing with videos, is gaining more and more popularity among contemporary scholars when it comes to studies of other modes. Practically, it is concerned with guiding effects in real CEC teaching practices to teachers, especially those novice ones.

Limited by the scale of the space of this thesis, only two cases are examined for a detailed and all-around analysis. More representative cases should be further analyzed to achieve a comprehensive picture by minimizing deviation.

Besides, in terms of modes, although this study has explored four major visual-verbal modes such as "image", "words", "dynamic", and "symbol" and has offered reasonable proof for selecting these four modes, the limitation lies in that other elements of visual-verbal modes, such as "color" or "size" are not fully studied here. Therefore, it is hoped that this study can arouse attention for further study as a brick casted to get a gem.

Acknowledgements

The author would like to express sincere thanks to the journal editor for reading this paper and further reviews.

References

- Baldry, A., & Thibault, P. J. (2006). *Multimodal Transcription and Text Analysis*. London: Equinox Publishing.
- Bezemer, C. (2008). Displaying Orientation in the Classroom: Students' Multimodal Response to Teacher Instructions. *Linguistics and Education*, 19(1), 166-178. <http://dx.doi.org/10.1016/j.linged.2008.05.005>
- Chen, X. R., & Q. Y. H. (2011). The Application of Multimodal Discourse Analysis in Pragmatic Research. *Foreign Languages in China*, 8(5), 89-93.
- Cloonan, A. (2011). Creating Multimodal Metalanguage with Teachers. *English Teaching: Practice and Critique*, 10(4), 23-40. Retrieved from <http://web.ebscohost.com>
- Cope, B., & Kalantzis, M. (2000). *Multiliteracies: Literacy Learning and the Design of Social Future*. London: Routledge.
- Halliday, M. A. K. (2001). *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.
- Hu, Z. L., & D. J. (2006). How Is Meaning Construed Multimodally—A case study of a PowerPoint presentation contest. *Computer-Assisted Foreign Language Education*, 109(6), 82-86.
- Kress, G., & van Leeuwen, T. (2001). *Multimodal Discourse: The Modes and Media of Contemporary Communication*. London: Arnold.
- Kress, G., & van Leeuwen, T. (2006). *Reading Images: The Grammar of Visual Design*. London: Routledge.
- Loretto, A., & Chisholm, J. (2012). Learning to Teach Generative Meaning-making through Multimodal Inquiry. *English Teaching: Practice and Critique*, 11(1), 136-149.
- O' Halloran, K. L. (2010). Systemic Functional-Multimodal Discourse Analysis (SF-MDA): Constructing Ideational Meaning Using Language and Visual Imagery. *Visual Communication*, (4), 443-475.
- Royce, T. (2002). Multimodality in the TESOL classroom: exploring visual-verbal synergy. *TESOL Quarterly*, 36(2), 191-205.
- Royer, T. (2007). Intersemiotic Complementarity: A Framework for Multimodal Discourse Analysis. In T. Royce & W. L. Bowcher (Eds.), *New Directions in the Analysis of Multimodal Discourse*. London: Lawrence Erlbaum Associates. Retrieved from <http://forlingua.com/downloads/Chap12Royce-Erlbaum.pdf>
- Schmid, E. C. (2008). Potential Pedagogical Benefits and Drawbacks of Multimedia Use in the English Language Classroom Equipped with Interactive Whiteboard Technology. *Computer and Education*, (51), 1553-1568. <http://dx.doi.org/10.1016/j.compedu.2008.02.005>
- The New London Group. (2000). A Pedagogy of Multiliteracies: Designing Social Futures. In B. Cope & M. Kalantzis (Eds.), *Multiliteracies: Literacy Learning and the Design of Social Future*. London: Routledge.
- Wei, Q. H. (2009). *Studies on Multimodality and Multimodal Discourse in Visual Surroundings*. Beijing: Science Press.
- Wen, Q. F. (2004). *Applied Linguistics: Research Methods and Thesis Writing*. Beijing: Foreign Language Teaching and Research Press.
- Zhang, D. L. (2010). Preliminary Investigation into the Concept of Design and the Selection of Modalities in Multimodal Foreign Language Teaching. *Foreign Languages in China*, 7(3), 48-53.
- Zhang, D. L., & Li, X. Y. (2012). Modality Synergy in Multimodal Classroom Discourse. *Foreign Languages and Their Teaching*, 262(1), 39-43.
- Zhang, L. X. (2012). A Study of Multi-modal Discourse Based on ELAN—A Case Study of a Teacher's Classroom Discourse of College English. *Modern Educational Technology*, 22(7), 54.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).