

# ELAN-Based Multimodal Analysis of Non-linguistic Features in the TED Speech for the Education of Public Speaking Ability

Yahui Chu<sup>1</sup>, Jing Chen<sup>2,\*</sup>

<sup>1</sup>*School of Foreign Languages, Hangzhou Normal University*

<sup>2</sup>*School of foreign languages, Hangzhou normal University*

\**Corresponding author. Email: jeanchan78@163.com*

## ABSTRACT

Non-linguistic features, conveying meaning and speech as symbolic resources, is an activity that involves a wealth of modes, where non-linguistic symbols can play an important role. Thus, under the guidance of dynamic multimodal analysis framework as well as the video analysis tool ELAN, an analysis of the speech by Han Xue, a famous actor and singer in China was carried out. Findings are overused movements with hands, using speech rhyme for attracting attention, as well as single facial expression. Some non-linguistic resources are contributing to meaning-making, yet Han fails to use them properly. Besides, when relations of modes in Han's speech are investigated, it turns out that highlighting relation is the most frequent one, and body movement is the least appropriately used resources. Regarding to the education of English speaking ability, learners might adopt appropriately non-linguistic features through multiple relations of modes.

**Keywords:** *ELAN, Multimodal analysis, Non-linguistic features, TED speech, Education of English speaking ability.*

## 1. INTRODUCTION

Multimodal discourse refers to communication involving various means of communication such as language, image, sound, and action as well as symbolic resources through the use of auditory, visual, tactile, and other senses. Hu (2007) suggested that all modes are on the tendency to form an interactive network in meaning. Zhu (2007) is one of the early authors who bring multimodal discourse analysis into domestic research practice, and probes into the theoretical basis and the procedures. Li (2003) firstly introduces in detail the visual grammar.

Theoretical researches on multimodality overreaches pragmatic ones, with pragmatic ones are virtually about the usage of theory to the study of image (Sun, 2012). Other symbolic resources except language are also vehicles of meaning, and the study of multimodal discourse contributes to the full understanding of discourse meaning (Guan, 2017; Wang, 2015).

This speech *To Be a Positive Pessimist* s given by Han Xue in 2016, which could be accessed by

<https://mp.weixin.qq.com/s/V8ZXSyPGHvxabzGEqAss9Q>. Han gives her life conviction that is to be a positive pessimist, which means to make the worst plan and try the best. Han Xue, as an actor and singer, impressed the audience a lot with her performance in TED because of her ability to give a speech.

This paper aims to address the following questions with the help of ELAN, which enables the user to underline the content, facial expressions, body gestures, voice, and tone of the speaker.

1. What are Han's characteristics of the speech rhyme, body language, and facial expression?
2. What are the relations of modes in Han's speech?

## 2. METHODOLOGY

### *2.1. Dynamic Multimodal Discourse Analysis Framework*

Based on Halliday's Systemic Functional Linguistics (1978), as well as Norris' work *Analyzing Multimodal Interaction: A Methodological Framework* (2004),

Zhang (2009b) proposed a synthetic framework for the analysis of multimodal discourse where there are four levels. Firstly, according to this framework, the culture level which is on the top of the framework plays a decisive role in fostering multimodal communication possible for prerequisites for communication. In a specific context, the factors in the situation context will restrict communication activities, namely field, tenor, and mode. These three concepts are from Systemic Functional Linguistics (Halliday, 1978). The third is the content level where there are two aspects: the semantic level and the form level. At the semantic level, there are three metafunctional meanings: ideational meaning, interpersonal meaning, and textual meaning. The form level is realized by the employment of different modes that can cooperate to establish a certain relationship to help construct the whole meaning of the discourse.

Zhang (2009a) divides the relationships between the lexico-grammar of different modes into complementary ones and non-complementary ones. Different modes are all alternatives for expressing the whole meaning of the speakers and application of them is the compensation of monomodal discourse's inadequacy in conveying the whole meaning of discourse. This kind of relationship between different modes is named as complementary. The complementary relationship can be further distinguished into two kinds, namely reinforcement, and non-reinforcement. The reinforcement relationship means that in a multimodal discourse, there is only one mode is the main one and the others are just aiming to highlight it. Non-reinforcement means all the modes in the multimodal discourse are equal to each other and they are indispensable and complementary. Non-complementary relation means the second mode has no greater significance than the first one but it is still used by the speaker as a kind of meaning-making mode.

## 2.2. Transcription Criteria of Dynamic Multimodal Discourse

Zhang (2009b) proposed criteria for segmenting the videos to conduct dynamic multimodal discourse analysis, namely grammar of image and discourse. According to the criteria, a story (discourse) can be divided into several plots, and a plot can be divided into many events. The events here are equal to the image. An image can be further segmented into a figure and then the figure can be divided into pixels. Besides, the discourse criteria for segmentation of a dynamic multimodal discourse as follows: (1) Discourse structure; (2) Empirical significance; (3) Foregrounding features and characteristics.

## 2.3. ELAN for Annotating Video

ELAN is a cross-platform multimodal discourse analysis software developed by Max Linguistic Institute of Psycholinguistics in the Netherlands.

ELAN enables users to define language types and tiers as well as annotate videos. Then annotations could be searched and replaced. Besides, the annotation could be saved as different files depending on the demands.

## 3. RESULT

This video could be divided into two parts: lead-in and the main content of speech. Here only the first part is under analysis and investigation. The discourse of the first part aims to lead in and concerns with the idea to be a positive pessimist. Transcription is illustrated in Figure 1. There are two episodes under this discourse. The first one is to share simple works by herself and the second one is to raise questions and give possible questions. And each episode is realized by three events. She aims to give examples of her own life so she puts forward three events for each episode. In the first, she plays a song that she used to sing when she was young. Followed by two pictures painted by her in her early childhood and adulthood. In the other episode, she intends to raise a question about the reason on the worse performance of her work, and she gives the possible answer, which is that adults may lack imagination. And the last event is the introduction of her strategy--to be a positive pessimist.

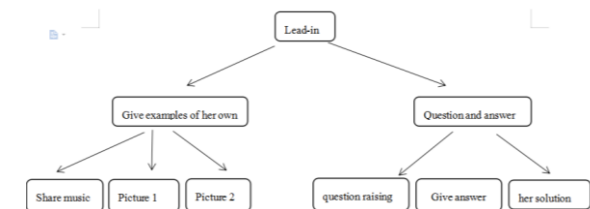


Figure 1 Transcription of the lead-in part

Focusing on the non-linguistic languages, there will be a deeper understanding of one's performance in speech when we try to figure the whole meaning conveyed by other social symbol resources.

### 3.1.1. Gestures

When talking about the gestures in the video, it can be noticed that most of the gestures are done by two hands. Only when in the beginning of the video on the PowerPoint does Han Xue slightly move her body to look at the screen. Then, the most frequently used gesture is "pointing to the screen" and "stretching hands out". "Pointing to the screen" is a kind of directive gesture, but "stretching hands out" does not have this kind of function. When "the pointing to the screen" gesture is taken into consideration, some differences could be also noticed on various occasions. When she

points to the screen, sometimes she would turn her face to the screen, and sometimes she would like to have eye contact with the audience. Besides, some gestures seem to have no relation with the words meaning that she tries to express. For example, when she says "so I know, you might have a question" she stretches her hands out, this kind of gesture does not match this visual modal and they are in a redundancy relation.

What's more, in her speech, there is a tendency to overuse gestures, "slightly point at" for instance. When she utters that "we dare not to try any mistakes", she points at every pause of a word, which shows her bad performance at using gestures to make the interpersonal relationship. She never uses direct gestures involving the audience to encourage them. For example, when she talks about the reason why we dare not to make mistakes, she just turns her palms outside without referring to the audience.

number	gestures in lead-in part	gestures
1	00:00:02.000 - 00:00:04.000	waving hands
2	00:00:05.747 - 00:00:06.843	point to screen
3	00:00:28.420 - 00:00:29.475	slightly turn body around
4	00:00:30.292 - 00:00:32.371	one hand out
5	00:00:33.000 - 00:00:33.755	slightly squeeze hands
6	00:00:34.755 - 00:00:36.818	one hand out
7	00:00:38.283 - 00:00:40.339	press remote button
8	00:00:43.315 - 00:00:48.040	point to screen
9	00:00:48.040 - 00:00:49.467	point herself
10	00:00:49.939 - 00:00:52.027	point to screen
11	00:00:54.595 - 00:00:55.707	press remote button
12	00:00:57.203 - 00:00:58.277	point to screen
13	00:00:59.280 - 00:01:01.359	point to screen
14	00:01:01.450 - 00:01:01.788	a gesture of two
15	00:01:04.858 - 00:01:05.939	two hands out
16	00:01:06.800 - 00:01:07.881	two hands out
17	00:01:08.380 - 00:01:10.461	two hands out
18	00:01:10.799 - 00:01:11.595	one hand out
19	00:01:14.560 - 00:01:16.000	one hand out
20	00:01:17.580 - 00:01:19.200	two hands out
21	00:01:22.964 - 00:01:24.012	turn palms outward
22	00:01:27.012 - 00:01:29.091	the right hand up and down
23	00:01:29.240 - 00:01:32.288	turn palms outward
24	00:01:34.100 - 00:01:37.148	turn palms outward
25	00:01:39.044 - 00:01:48.907	slightly point
26	00:01:49.900 - 00:01:52.948	two hands up and down
27	00:01:53.040 - 00:01:56.088	turn palms outward
28	00:01:56.180 - 00:01:56.500	turn palms outward
29	00:01:56.880 - 00:01:57.928	turn palms outward
30	00:02:00.120 - 00:02:02.168	right hand up
31	00:02:02.980 - 00:02:04.628	left hand up
32	00:02:04.880 - 00:02:05.928	press remote button
33	00:02:07.200 - 00:02:10.248	right hand up and down

Figure 2 Gestures in the lead-in

3.1.2. Intonation

Intonation, as one of the suprasegmental features of utterance, is also considered here because of its importance in English utterance. It is this utterance marked in the between of falling tone and raising tone, or level tone and raising tone that entails further information. Those comparatively huge changes of intonation are emphasized in the picture. Her usage of raising tone is mainly for attracting attention (Number 5 and 6; 13 and 14; 16 and 17; 23 and 24) or for delivering important ideas such as in Utterance 34 and 35 (See Figure 4).

WORDS	number	time	intonation
Good afternoon, everyone.	1	00:00:02.489 - 00:00:04.758	F1L
First let me share a piece of music with	2	00:00:05.234 - 00:00:07.882	F1F
I know it's terrible	3	00:00:32.088 - 00:00:33.055	F1F
because I played it	4	00:00:33.494 - 00:00:34.927	F1F
and I do know it's not the worst one	5	00:00:38.200 - 00:00:40.590	F1F
I still got something to show you	6	00:00:39.200 - 00:00:40.590	F1F
Don't I teach	7	00:00:41.880 - 00:00:43.025	F1L
it's an entire story	8	00:00:43.025 - 00:00:44.620	F1L
if you really want to understand	9	00:00:44.620 - 00:00:46.260	F1L
what is it talking about	10	00:00:46.260 - 00:00:47.660	F1L
you can check my Weibo later	11	00:00:47.660 - 00:00:49.000	F1L
it is this picture I painted last year	12	00:00:51.791 - 00:00:56.471	F1F
I have another one to show you	13	00:00:56.471 - 00:00:57.000	F1F
now	14	00:00:58.860 - 00:00:57.000	F1F
but	15	00:00:57.340 - 00:00:59.279	F1L
it seems better now	16	00:00:59.400 - 00:01:00.000	F1F
this picture	17	00:00:59.400 - 00:01:00.760	F1F
I painted	18	00:01:00.880 - 00:01:02.640	F1F
20 years ago	19	00:01:02.240 - 00:01:03.743	F1L
so I know	20	00:01:03.000 - 00:01:03.555	F1L
you might have a question	21	00:01:05.360 - 00:01:07.580	F1L
how it's getting worse	22	00:01:08.300 - 00:01:10.151	F1L
and I have understood that most of your	23	00:01:11.119 - 00:01:14.600	F1F
conclusion must be	24	00:01:14.600 - 00:01:16.515	F1F
out of imagination	25	00:01:18.008 - 00:01:19.421	F1F
But how it happened	26	00:01:20.825 - 00:01:21.840	F1F
I think it's because	27	00:01:21.640 - 00:01:23.087	F1F
I became an adult	28	00:01:23.367 - 00:01:23.169	F1F
since we became adults	29	00:01:26.791 - 00:01:29.040	F1L
we started being afraid of making	30	00:01:29.440 - 00:01:33.555	F1L
the choice to be the best	31	00:01:34.263 - 00:01:38.340	F1F
or nothing	32	00:01:38.340 - 00:01:39.400	F1F
we do not dare to try any mistakes.	33	00:01:39.400 - 00:01:37.420	F1F
It seems that we trapped ourselves	34	00:01:38.300 - 00:01:39.247	F1F
in this over-competitive world	35	00:01:40.000 - 00:01:42.000	F1F
And since then	36	00:01:43.824 - 00:01:47.220	F1L
we started to lose a lot of abilities	37	00:01:47.220 - 00:01:48.904	F1L
like	38	00:01:50.028 - 00:01:54.907	F1L
imagination	39	00:01:55.100 - 00:01:55.878	F1L
and creativity	40	00:01:55.911 - 00:01:58.491	F1F
what is the situation	41	00:01:57.000 - 00:01:57.668	F1F
For me I have a special one	42	00:02:00.388 - 00:02:02.180	F1F
(it) is being a positive pessimist.	44	00:02:07.271 - 00:02:10.423	F1F

Figure 3 List of intonation and sentences

3.1.3. Facial Expression

Facial expression is also a vital tool in conveying meaning. When someone compliments you, you would more likely to accept that when she smiles at you rather than a seriously unhappy look. After analyzing the facial expression in Han's video, the authors find that in most of the speech, she smiles. When raising questions, important ideas, she would make a serious look, but then her look would immediately change into a smiling face. Her facial expression helps to convey meaning to some extent.

When showing her drawing as an example, her embarrassing look on her face is complementary to the picture in the PowerPoint, which makes the audience burst into laughter. Another little embarrassing smile also occurs when she gives another picture of herself (See Figure 4). Above all, the authors might have a conclusion that Han fails to use facial expression to convey meaning, or at least, she is not good at using the facial expression as a meaning potential resource.

words	First let me share a piece of music with you
TC	00:00:05.234 - 00:00:07.882
SD	(22.91)
words	emm
TC	00:00:30.800 - 00:00:32.088
facial expression	little embarrassing smile
TC	00:00:31.865 - 00:00:34.745
words	I know it's terrible
TC	00:00:32.088 - 00:00:33.055
SD	(0.43)
words	because I played it
TC	00:00:33.494 - 00:00:34.927
words	I have another one to show you
TC	00:00:54.864 - 00:00:55.783
SD	(0.07)
words	now
TC	00:00:55.860 - 00:00:57.000
SD	(0.33)
facial expression	little embarrassing smile
TC	00:00:57.330 - 00:01:00.160

Figure 4 Han's embarrassing smile

## 4. MULTIMODAL ANALYSIS OF THE SPEECH

### 4.1. Cultural Context

Cultural context is the key aspect that makes multimodal communication possible, including the ideology composed of people's mode of thinking, philosophy of life, habits and all social unspoken rules, and the communicative procedure or structural potential. Cultural context determines the different communicative stages and steps that people adopt according to different communicative purposes, namely genres.

When in the leading part of a speech, the orator has to arouse the audience's interest in what he or she is going to talk about that is what the lead-in part for. Normally, there are self-introduction, warm-up, and then questions or information that he or she will focus on in a later speech. So Han also observes this routine. It begins with her self-introduction, warming up period by using two pictures in her early life, and then the question of why the earlier picture looks better than the later one. She brings up her answer that is because of the lack of imagination. Finally, she puts forward her solution to this problem, i.e. to be a positive pessimist.

### 4.2. Context of Situation

Context of the situation is the concrete embodiment of cultural context, including three variables: field which refers to social activities involved in discourse; tenor which is the social relationship between communicative parties and mode which is the medium and channel of communication. In this speech, the analysis of the context of the situation are as follows:

**Field:** It is a speech by Han, who aims to inform her audience about her life conviction, that is to be a positive pessimist.

**Tenor:** Two parties are in the relationship between an orator and the audience.

**Mode:** The interaction between the orator and audience is virtually a one-way activity. The orator is kind and also serious, dressing up decently. She uses simple words, soft sound, and relaxing intonation. Besides, she chooses modes formed by the continuous changes of her body and non-body media (background picture) as the visual mode to spread the image meaning of her speech.

### 4.3. Relations of Modes in Multimodal Discourse

#### 4.3.1. Overall Relation of Modes

Collaboration between the two modes is both necessary and inevitable. In the speech, there are mainly

three different modes. They are the auditory mode mainly formed by the lecturer's words, the visual mode formed by her body movement, and the visual mode formed by the screen, to be more specific, mainly the pictures on the screen.

Mode relation and those modes involved in are annotated by researcher. Results are listed in the picture below. Here we can once again notice some interesting points. Firstly, it can be found that there are two periods where no relation of modes is marked out. And the total lasting time is almost 37 seconds, which is at a proportion of 28%. It means that she only uses a single auditory mode to express her idea and this counts one-quarter of the time. In this circumstance, she might want to draw the audience's attention to her words but this single mode use is likely to yield boredom. According to the table above, she uses body movement as well as other modes, but they can not compose the whole meaning, we can tentatively conclude that she did not use these resources well. Then we can find out that her hand movement is least engaged in multimodality use. What's more, highlighting relation is the most frequent one in terms of the times it occurs. Highlighting does help to emphasize an idea, which is vital in the main content part of speech but are not supposed to be frequently used in the lead-in part. The overuse of multimodalities in the speech causes it is lengthy and inefficient.

relation of modes	modes involved	timeline	lasting time(seconds)
highlighting	smile&words	00:00:00.281 - 00:00:03.750	5.969
Intersection	music&facial expression	00:00:05.750 - 00:00:36.500	30.750
primary and second relation	picture&words	00:00:36.500 - 00:00:49.800	13.300
coordination	picture&words	00:00:49.800 - 00:01:08.443	18.643
no	no	00:01:08.443 - 00:01:25.139	16.696
extension relation	facial expression&words	00:01:25.139 - 00:01:33.491	7.283
highlighting	hand movement&Words	00:01:33.491 - 00:01:37.164	3.673
no	no	00:01:37.164 - 00:01:58.259	20.099
highlighting	hand&words	00:01:58.259 - 00:02:10.363	12.104

**Figure 5** Relation of modes

A different relationship between these modes contributes to the whole meaning. So, in this part, the authors intended to analyze their coordination. The visual mode has two main functions on auditory mode. One is strengthening, in which the audience can infer from visual mode to make the information obtained through auditory mode clearer and more accurate. And the other is supplementing or compensating, which is to supplement or compensate for the information missing in auditory modes or not clearly communicated or not fully received by the audience. After analyzing the relationship between different modes in Han's speech, there is some typical relation in composing meaning.

#### 4.3.2. Reinforcement

When in the beginning, the music composed by herself is playing to create an atmosphere of serenity, and her relaxing facial look, as well as relaxed body, helps to compose this atmosphere. So this auditory mode and visual mode are in the reinforcement relation.

Her facial look helps reinforce the meaning of the music.

Then when she wants to comment on her piece of music, she turned her face into a smile, which once again is a reinforcement to her words, i.e to the auditory mode.

When she brings out the question, her face turns into a confusing look. Here she uses the visual mode to highlight the meaning made by the words, it is in a relation of primary and secondary. When the question "Why is the second work much better than the first" triggers audience's thinking, the facial expression of Han helps to lead the audience to focus on this question. Thus, the auditory mode is of the primary position and the visual one the secondary.

#### 4.3.3. *Non-reinforcement Relation*

She displays the pictures she made in her younger days. She gives information by speaking as well as showing the picture on the screen. The information she utters is the auditory mode, and the picture displayed on the screen is the visual stimulus and they are in the relation of coordination, which means these two modes construct the whole meaning of the speaker and the lack of one mode will make the meaning of discourse uncompleted.

### 5. CONCLUSION

To bring the paper to a close, main points are summarized here. Though Han Xue's English mastery is far above the average level, her speech given on TED is still of some inadequacies. She overuses her hand's movements and some of them are of no help to the meaning-making of the discourse. Besides her single facial expression and level rhyme indicates that she fails to use these symbolic resources properly. There might be cultural and societal factors. For instance, Chinese people tend to be moderate, while English learners in China are evaluated mainly by their linguistic competence rather than their linguistic performance. Additionally, Zhang (2009b) posited that the main topic of multimodal discourse research is to find out how different modes are related to each other to achieve the intention of the speaker. Thus, the relations of different modes are also investigated, in which highlighting is the most frequent one, making the lead-in part lengthy.

As we have seen above, it is not enough for a lecturer with a well mastery of linguistic knowledge to deliver a sound speech. Both teachers and students have to be aware of the significant role and cooperation patterns of non-linguistic features.

### ACKNOWLEDGMENTS

This work is supported by the Ministry of Education Research of Social Sciences Youth funded projects, China (Grant No. 17YJC740009).

### REFERENCES

- [1] Halliday, M.A.K, *Language as a Social Semiotic*. London: Edward Arnold, 1978.
- [2] Norris, S. *Analyzing Multimodal Interaction: A Methodological Framework*. London: Routledge, 2004.
- [3] Guan, Y. 'Dongtai yanjiang shipin de duomotai fenxi' [Multimodal discourse analysis of dynamic speech video], *Zhejiang ligong daxue xuebao (shehui kexue ban)* [Journal of Zhejiang Sci-Tech University (Social Sciences)], 2017(1), pp. 25-32.
- [4] Hu, Z.l. 'Shehui fuhaoxue yanjiu zhong de duomotaihua' [Multimodalization in social semiotics], *Yuyan jiaoxue yu yanjiu* [Language Teaching and Linguistic Studies], 2007(1), pp. 1-10.
- [5] Li, Z.z. 'Duomotai fuhaoxue lilun jichu yanjiu tujing yu fazhan qianjing' [Multimodal semiotics: theoretic basis, research approach, and development prospect], *Waiyu yanjiu* [Foreign Languages Research], 2012(1), pp. 1-8.
- [6] Sun, Y. 'Duomotai huayu yiyi jiangou yi 2011 xi an shijie yuanyi bolanhui huihui wei jidian' [The meaning construction of multimodal discourse: the emblem of 2011 Xi'an International Horticultural Exposition as an example], *Waiyu yu waiyu jiaoxue* [Foreign Language and Foreign Language Teaching], 2012(1), pp. 44-47.
- [7] Wang, Y.f. 'Duomotai huayu fenxi shijiao xia de yingyu yanjiang jiqiao' [An Analysis of English Speech Skills from Multimodal Discourse Analysis Perspective], *Shanxi datong daxue xuebao* [Journal of Shanxi Datong University], 29(2), (2015), pp. 90-92.
- [8] Zhang D.l. 'Duomotai huyu lilun yu meiti jishu zai waiyu jiaoxue zhong de yingyong' [Application of Multimodal Discourse Theory and Media Technology in Foreign Language Teaching], *Waiyu jiaoxue* [Foreign Language Education], 30(4), (2009a), pp. 15-20.
- [9] Zhang D.l. 'Duomotai huayu fenxi zonghe lilun kuangjia tansuo' [On A Synthetic Theoretical Framework for Multimodal Discourse Analysis], *Zhongguo waiyu* [Foreign Languages in China], 6(1), (2009b), pp. 24-30.
- [10] Zhu, Y.s. 'duomotai huayu fenxi de lilun jichu yu yanjiu fangfa' [Theory and Methodology of Multimodal Discourse Analysis], *Waiyu xuekan* [Foreign Language Research], (2007)5, pp. 82-86.