

# A Multimodal Discourse Analysis of TED-Ed Medical Popular Science Videos

Huichao Wu

Ningbo Tech University, Ningbo, China

Email: 1441306903@qq.com

**How to cite this paper:** Wu, H. C. (2025). A Multimodal Discourse Analysis of TED-Ed Medical Popular Science Videos. *Open Journal of Social Sciences*, 13, 340-357.  
<https://doi.org/10.4236/jss.2025.137020>

**Received:** May 12, 2025

**Accepted:** July 15, 2025

**Published:** July 18, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

In medical popular science communication, the dissemination of knowledge more and more employs multimodal discourse instead of just relying on textual descriptions and verbal explanations. However, the research in medical popular science videos has long been overlooked. This research selects the medical science video series titled *Know Your Body* from TED-Ed platform as the research subject, which has 12 episodes. These videos cover a wide range of topics with strong scientific validity, encompassing multiple disciplines such as physiology, immunology and disease prevention. They are presented in a multimodal manner, combining animation, narration, text, and sound effects, and are typical representatives of TED-Ed's multimodal science communication. Based on Zhang Delu's multimodal discourse analysis theory, this research uses the Elan software to analyze the multimodal discourse construction of these videos. The study finds that the relationship between animation, language, and images is mainly complementary, with language conveying core information and images serving as supplements or emphasis. Complementary-non-enhancement relationship and non-complementary relationship are not common in these videos. They usually play a transitional role, helping audience to make associations and think more comprehensively. The research findings can give theoretical and practical guidance in making multimodal videos for medical science popularization.

## Keywords

Medical Popular Science, Multimodal Discourse, Systemic Functional Linguistics, Health Communication

## 1. Introduction

In recent years, with people paying more attention to health and medical science advances rapidly, the need for medical science popularization is growing. China

has made many relevant policies one after another, aiming to make science popularization work better and spread wider. For example, in June 2024, the Central Committee of the CPC and the State Council together put out the *Opinions on Further Strengthening the Work of Science and Technology Popularization in the New Era*.

Traditional medical popular science content mainly relies on books, magazines, newspapers, television, etc. However, in today's increasingly convenient access to information, especially with the rise of new media, short videos have become a new medical popular science communication carrier. Short videos not only have the characteristics of rapid dissemination, large amount of information and strong entertainment, but also attract the audience's attention through vivid visual impact, so as to effectively improve the public's cognition and awareness of medical knowledge.

Within this context, the concept of multimodal discourse construction has gradually attracted attention. It emphasizes conveying information through different means (such as images, words, sounds, etc.). Compared with the more serious forms of popular science communication in the past, the videos on social media usually combine animation, live-action footage, diagrams, dubbing and other forms together, so that complex medical knowledge is presented to the audience in a more understandable way. Being more conducive to memory retention, such videos can enhance the acceptability of information, and promote the public scientific literacy.

Within the domain of multimodal discourse research, [Kress and Leeuwen \(1996\)](#) defined images as social symbols, developing a social semiotic approach based on visual grammar. Later, [Norris \(2004\)](#), who, in her work *Multimodal Interaction Analysis: A Methodological Framework*, proposed a varied framework for multimodal interaction analysis.

Meanwhile, many scholars tried different approaches to multimodal discourse analysis. [Forceville \(1996, 2009\)](#) pioneered multimodal metaphor research. [Li \(2003\)](#), [Zhu \(2007\)](#) and [Xin \(2008\)](#) also contributed many new perspectives on multimodal discourse construction, pushing forward its academic development. [Liu & Hu \(2015\)](#) explored corpus-based multimodal discourse analysis. [Zhang & Zhang \(2022\)](#) combined multimodal critical discourse analysis and multimodal positive discourse analysis theories to propose an integrated framework based on the production and interpretation processes of multimodal texts.

In practical discourse analysis, [Wang & Fan \(2019\)](#) did a full study of about 289 health communication accounts. [Alhassan et al. \(2019\)](#) did a study by using short videos to teach Turkish students about sexually transmitted diseases, showing these videos can be a good and cheap way for young people's health education. [Liu & Zhao \(2020\)](#) combined multimodal data with face-to-face interviews to see how short videos help to spread health knowledge. [Romano et al. \(2022\)](#) looked at 210 entries and proposed that YouTube can be a good platform for sharing medical science content quickly. [Tereszkiewicz \(2023\)](#) studied how Polish doctors use different video methods to build special interactive styles with audience.

Based on the above previous study, this research intends to conduct an in-depth study on the multimodal discourse construction of short medical science popularization videos, especially analyzing how short videos integrate visual and linguistic modalities to effectively disseminate scientific information and create works with more communication power. This not only has important theoretical significance, but also provides practical reference for improving the communication quality of medical science popularization content.

## 2. Theoretical Framework

Zhang (2009) put forward a whole theory system for multimodal discourse analysis, which is based on systemic functional linguistics and pays attention to study the forms and relations of multimodal discourse from different angles. This system includes five main parts: cultural part, situational part, semantic part, formal part, and media part. The research mainly looks at how different forms of multimodal discourse connect with each other in the formal part. Zhang Delu categorizes the relationships between the forms of multimodal discourse into two major types: complementary relationships and non-complementary relationships, as shown in Figure 1.

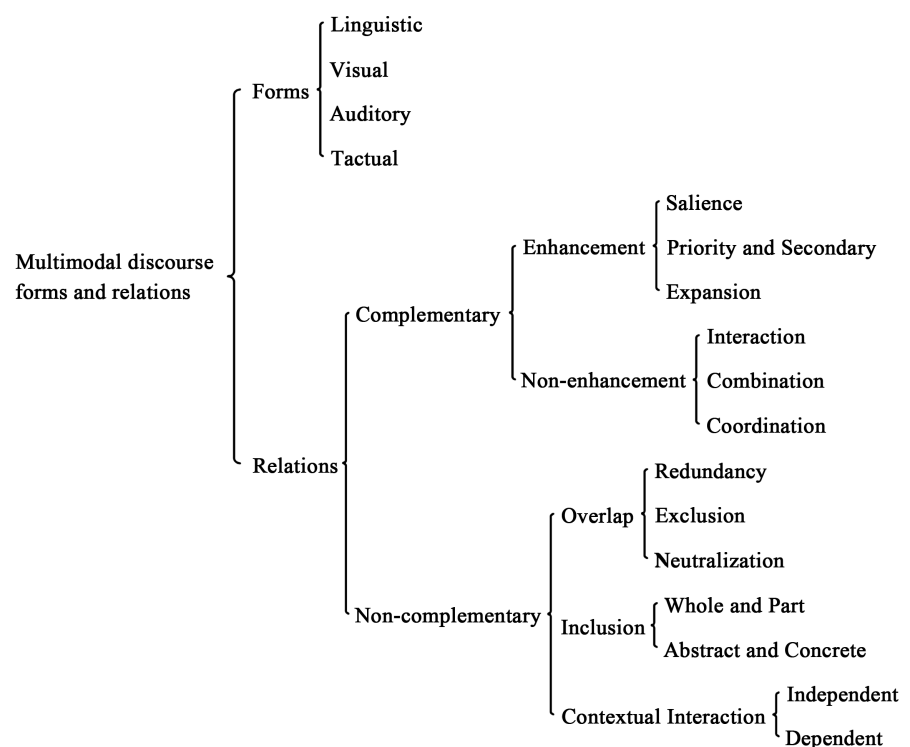


Figure 1. Multimodal discourse forms and relations from Zhang (2009).

### 2.1. Complementary Relation

In complementary relationships, images and texts work together to pass information with their combined effect. Here, one way becomes the main communi-

---

cation form while the other plays a supporting role. There are two types of complementary relations: enhancement and non-enhancement.

### **2.1.1. Enhancement**

Enhancement relations can be divided into three kinds: salience, priority-secondary, and expansion. In salience relation, pictures may catch people's eyes by making them bigger or changing colors. Meanwhile, words help to express the same information in a clear and simple way. For priority-secondary relation, pictures and words express information in a way that is commensurate with their respective importance. Expansion relation means pictures and words work together to develop the same topic, with words usually giving more details or background to make content richer.

### **2.1.2. Non-Enhancement**

Non-enhancement relationships are divided into three types: interaction, combination, and coordination. In interaction relationships, images and text are both necessary and independently convey their respective information. They jointly express the same concept or theme but do not rely on each other. In combination relationships, different types of media of the same mode work together to convey meaning. For example, the background music and the voice of narration can combine with each other to convey the ideas. In coordination relationships, the two modes intersect, presenting different content to make the overall meaning.

## **2.2. Non-Complementary Relation**

Non-complementary relationships may not seem as important as the complementary category, but actually they are indispensable. This kind of relationship usually comes in three types: overlap, inclusion, and contextual interaction. Overlap means two or more modes appear together but the additional modes don't contribute to new meanings. Inclusion means one or more modes, either same or different, give more details but not really new meanings. Contextual interaction is where context can join in the communication but do not take active part; instead, it is pulled into the communication process based on the communicator's goals and the adopted communicative methods.

### **2.2.1. Overlap**

In overlap relationships, sometimes images and texts provide the same information, serving the viewer with repeated content, which is referred to as redundancy. However, this kind of redundancy is positive and can often bring convenience to the audience. But there is another kind of overlap where images and texts contradict with each other, leading to confusion for the viewer. Noise is a typical example. This is known as the phenomenon of exclusion and neutralization, and such overlaps are often passive or undesirable.

### **2.2.2. Inclusion**

Inclusion can be divided into two types: whole-part, and abstract-concrete. In the

whole-part relationship, the second mode typically provides part of the information rather than adding entirely new information. In the abstract-concrete relationship, although the second mode does not provide different information, it makes the information from the first mode more specific and concrete.

### 2.2.3. Contextual Interaction

Contextual interaction relationships are divided into contextual independence and contextual dependence. This is reflected in whether the viewer's environment is related to the modes being encountered. If environmental factors do not participate in the discourse communication, it shows contextual independence. However, if environmental factors participate in the overall meaning expressed by the discourse, the discourse will exhibit strong contextual dependence.

## 3. Research Method

### 3.1. Research Object

This research selected 12 medical popular science videos under the series title of *Know Your Body*, which was released by TED-Ed between 2012 and 2019 as the research object:

- 1) *Does stress cause pimples?*
- 2) *What do the lungs do?*
- 3) *How does the thyroid manage your metabolism?*
- 4) *How to grow a bone?*
- 5) *What would happen if you didn't sleep?*
- 6) *Why do people have seasonal allergies?*
- 7) *Why do you need to get a flu shot every year?*
- 8) *How does your immune system work?*
- 9) *What happens during a stroke?*
- 10) *How did teeth evolve?*
- 11) *What causes body odor?*
- 12) *Your body vs. implants*

The videos from 2012 to 2016 mainly use simple visual and sound parts, showing content in a direct and easy way. The modal interactions in these videos are quite basic, focusing on teaching knowledge through clear pictures and sounds. For example, *What do the lungs do?* uses animation to show how lungs work, with the auditory narration directly matching the visuals, making the information delivery stronger for viewers. On the other hand, the videos from 2017 to 2019 show a clear change, often starting with a short quote serving as a thematic thread throughout the video. These newer videos not only have richer visuals, sounds, and animations but also include multimodal information like body language. The combination of these modes creates a more complex and detailed way of presenting knowledge. For instance, *How does your immune system work?* and *Your body vs. implants* use advanced animation and sound mixing to make hard biological ideas easier to understand, while adding text and body cues helps viewers

grasp the content better. These elements work together in the videos, forming a complete and logical structure.

### 3.2. Annotation and Statistical Method

The research uses ELAN 6.8 video analysis software (<https://archive.mpi.nl/ta/elan>) as the annotation tool. It is a tool specially made for multimodal discourse study. It enables precise temporal alignment down to the 0.01-second level, and it can cut, mark, locate and loop play video materials. Also it supports multi-layer annotation, making it possible to analyze pictures, sounds, texts and other modes together.

In this research, text-image relationship means how words (like narrations and explanations) and images (including animations and pictures) work together in multimodal texts.

TED-Ed medical science videos usually last about 5 minutes and they have many different kinds of multimodal symbols inside. To facilitate the quantitative analysis of dynamic discourse, we annotate the corpus from six aspects: language, animation, picture, the relationships between language and animation, language and picture, and animation and picture, as shown in **Table 1**.

Language, being the most basic way for people to communicate, mainly helps to express ideas, give reasons, and tell stories. In TED-Ed medical science videos, language often works with moving pictures or drawings to better convey exact medical facts and background knowledge. The mode of language includes the narration that moves the story forward and the explanation that explains science ideas in detail. Narration usually gives the setting or background and often comes at the start or end of the video to bring up the subject, set the tone or help with transition. Objective narration keeps a fair tone, while subjective narration carries stronger emotions. On the other hand, explanation means the clear instruction of science rules and knowledge, often by a speaker, mainly to help people understand the hard or special information.

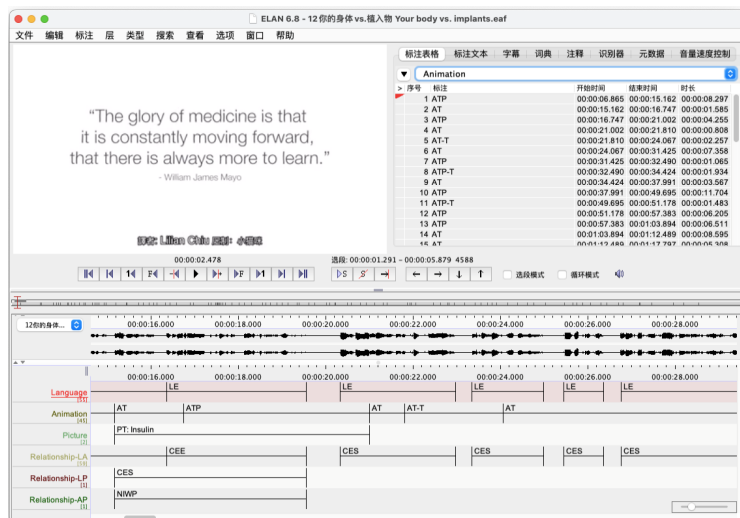
As a dynamic visual form, animation can intuitively present complex biological and medical processes. TED-Ed videos mainly rely on animation to depict the inner workings of cells, organs, and other living organisms, making abstract concepts more vivid and helping the audience better understand and retain the information. Animation includes dynamic visuals, such as representations of objects (“things”), human figures (“people”), or a combination of both (“people & things”), sometimes accompanied with dynamic texts.

As a static visual element, pictures are typically used to complement details that text or animation cannot convey. Although the use of pictures is relatively limited in TED-Ed videos, they still play an important role in highlighting key information and reinforcing visual memory. Picture refers to static images, which can either highlight specific words (“highlighted words”) or provide detailed views of concepts or objects (“close-up pictures”).

Based on **Table 1**, the videos were annotated in layers. The segmented annotations can be repeatedly reviewed and modified. Additionally, statistical results can be exported for quantitative analysis, as illustrated in **Figure 2**.

**Table 1.** Coding scheme.

Type	Coding	Meaning
Language	LN	Language-Narration
	LE	Language-Explanation
Animation	AT(-T)	Animation-Things(-Text)
	AP(-T)	Animation-People(-Text)
	ATP(-T)	Animation-Things & People(-Text)
Picture	PHW	Picture-Highlighted Words
	PCP	Picture-Close-up Pictures
	PWP	Picture- Words & Picture
	CES	Complementary-Enhancement-Saliency
	CEPS	Complementary-Enhancement-Priority & Secondary
	CEE	Complementary-Enhancement-Expansion
	CNI	Complementary-Non-enhancement-Interaction
	CNCom	Complementary-Non-enhancement-Combination
	CNCoo	Complementary-Non-enhancement-Coordination
	Relationship-LA (Language & Animation)	NOR
Relationship-LP (Language & Picture)	NOE	Non-complementary-Overlap-Exclusion
Relationship-AP (Animation & Picture)	NON	Non-complementary-Overlap-Neutralization
	NIWP	Non-complementary-Inclusion-Whole & Part
	NIAC	Non-complementary-Inclusion-Abstract & Concrete
	NCII	Non-complementary-Contextual Interaction-Independent
	NCID	Non-complementary-Contextual Interaction-Dependent



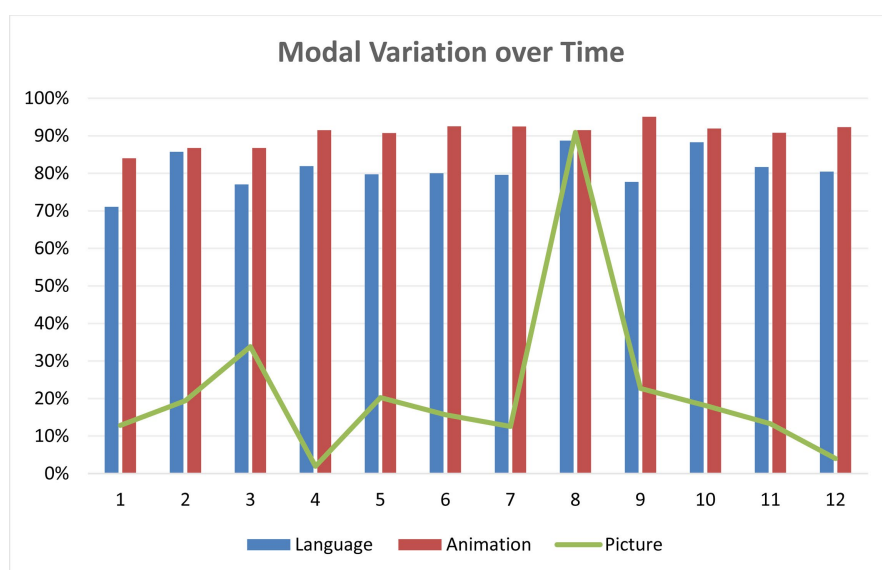
**Figure 2.** Elan hierarchical annotation diagram.

## 4. Analysis of Text-Image Relations in TED-Ed Medical Popular Science Videos

### 4.1. Overall Distribution of Three Modes

In TED-Ed medical science videos, the formation and expression of text-image relationships rely on different modalities (language, animation, and picture). These modalities collaborate in distinct ways to convey information, collectively forming a complete multimodal discourse structure. Understanding the distribution and variation of these modalities not only helps reveal how video content is presented but also aids in analyzing their roles in information transmission.

Based on the data generated from ELAN annotation, we can observe variations in the proportions of language, animation, and picture across multiple TED-Ed videos. Below is the overall distribution of these three modalities:



**Figure 3.** Modal data analysis.

**Figure 3** shows that animation consistently occupies the largest proportion of time, followed by language, while pictures account for the smallest share. Animation keeps high percentage all the time, with an average of 90% across different years. The extensive use of animation facilitates a more intuitive and clear demonstration of dynamic processes, such as biological changes inside the human body, thus enhancing the audience's learning experience. Besides, the dominance of animation in TED-Ed videos significantly contributes to viewers' comprehension and memory. Animated sequences allow complex medical processes (e.g., blood flow, cellular mechanisms) to be visualized dynamically, thereby enhancing mental models. Research in cognitive theory supports this, suggesting that dynamic visuals aligned with verbal explanations improve understanding more effectively than static text or image alone.

Language also occupies a high proportion, with an average of about 80%. Language plays a crucial role in explaining complex medical ideas. Picture accounts

for the smallest share, maintaining a relatively low level across most years, with an average of 16%. As a static element, pictures primarily serve a supplementary function.

#### 4.2. Overall Distribution of Complementary and Non-Complementary Relationships

After using the ELAN software to annotate the 12 TED-Ed medical science videos and to statistically analyze the complementary and non-complementary relationships among language, animation, and pictures, the results are presented in **Table 2** and **Table 3**.

Data in **Table 2** reveals the complementary relationships among language, animation, and pictures in TED-Ed medical science videos. To save space in **Table 2**, the list of videos only uses key words in the episode titles. Based on Zhang Delu's multimodal discourse framework, the analysis highlights distinct modality combinations and integration strategies across different medical topics.

**Table 2.** Proportion distribution of complementary relationships among language, animation and pictures<sup>1</sup>.

Video	Relationship-LA	Relationship-LP	Relationship-AP
1) Stress & Pimples	66%	8%	3%
2) Lungs	72%	18%	0%
3) Thyroid	80%	33%	2%
4) Grow a Bone	84%	6%	0%
5) Sleep Deprivation	89%	13%	0%
6) Seasonal Allergies	84%	15%	0%
7) Flu Shot	92%	8%	0%
8) Immune System	89%	18%	0%
9) Stroke	93%	21%	0%
10) Teeth Evolution	91%	17%	0%
11) Body Odor	88%	10%	0%
12) Body vs. Implants	78%	2%	0%

According to **Table 2**, Language-Animation (LA) is the primary complementary mode. In all the videos, the proportion of Language-Animation was significantly higher than the other two types of relationships. The lowest is 66% (video

<sup>1</sup>LA is the abbreviation of "language" and "animation", LP is the abbreviation of "language" and "picture", AP is the abbreviation of "animation" and "picture". During the annotation process, there are cases where a video only has one modality (for instance, it has only animation modality but no language modality). In such cases, there is no relationship among the modalities, and thus the total sum of the data does not reach 100%.

1), and the highest is 93% (video 9), with overall proportion remains above 60%. This indicates that TED-Ed has established a relatively mature structure of “language + animation”. Such multimodal coordination not only reinforces viewer comprehension but also sustains engagement across different age groups. Videos focusing on physiological mechanisms (e.g., *What Happens During a Stroke*) tend to exhibit a highly complementary Language-Animation relationship.

In **Table 2**, complementary relationship between language and picture (LP) fluctuates between 33% (video 3) and 2% (video 12). Because in complex topics, the explanation of an idea may need more stationary pictures to help people understand. Complementary relationship between animation and pictures (AP) is 0% in most videos, possibly because of the dominant role of animation in visual storytelling, which leaves limited space for simultaneous appear of static pictures. But in rare cases (video 1 and 3) animation goes on while an important picture remains on the screen for a while so as to emphasize the point.

For non-complementary relationships, **Table 3** exhibits the general distribution. Since language and animation are generally designed to arrange together in popular science videos, the proportion of non-complementary LA relationship is generally very low, but LP and AP relationship shows more complicated results.

**Table 3.** Proportion distribution of non-complementary relationships among language, animation and pictures<sup>2</sup>.

Video	Relationship-LA	Relationship-LP	Relationship-AP
1) Stress & Pimples	1%	2.37%	4%
2) Lungs	1%	9%	13%
3) Thyroid	1%	20%	31%
4) Grow a Bone	2%	2%	2%
5) Sleep Deprivation	1%	8%	11%
6) Seasonal Allergies	2%	3%	15%
7) Flu Shot	2%	2%	7%
8) Immune System	2%	18%	18%
9) Stroke	1%	7%	19%
10) Teeth Evolution	1%	2%	15%
11) Body Odor	3%	9%	8%
12) Body vs. Implants	2%	2%	2%

According to **Table 3**, the non-complementary relationship between language

<sup>2</sup>During the annotation process, there are cases where a video only has one modality (for instance, it has only animation modality but no language modality). In such cases, there is no relationship among the modalities, and thus the total sum of the data does not reach 100%.

and pictures (LP) generally remains low but sometimes goes up as in video 3 and 8. When language and picture do not complement each other, there might be visual breaks, letting people pause or change focus. These still pictures, often close-up shots or key words, may help people think, understand new things or remember old ones. So even when pictures do not closely follow speaking, this arrangement can still better the watching experience by giving people time to think and digest.

The percentage of Animation-Picture (AP) non-complementary relationship is not very high in general, but when they do occur, they might reflect an intentional design. For instance, while animation conveys biological functions, a static diagram or metaphorical picture may offer a new interpretive angle, leading viewers to combine different modes to have a deeper understanding.

### 4.3. The Detailed Analysis of Complementary Relationships

Through video analysis and annotation, the three relationships—Language-Animation (LA), Language-Picture (LP), and Animation-Picture (AP)—exhibit distinct distribution patterns in TED-Ed medical science videos, reflecting how different modes collaborate. We extract percentage data for the six categories CES, CEE, CEPS, CNI, CNCom, and CNCoo<sup>3</sup> from the annotated tables and calculate their average proportions in Relationship-LA, Relationship-LP, and Relationship-AP. The processed results are presented below in **Table 4**, followed by specific examples for further analysis.

**Table 4.** The average distribution of complementary relationships.

Category	Relationship-LA	Relationship-LP	Relationship-AP
CES	64.12	9.23	2.64
CEE	14.23	5.34	0
CEPS	7.12	1.55	0
CNI	0	0	0
CNCom	0	0	0
CNCoo	0	0	0

#### 4.3.1. Relationship-LA (Language-Animation)

Complementary-Enhancement-Salience (CES, 64.12%) is the primary complementary relationship type in LA relationship, indicating that in most cases, animation and language form an enhancement relationship, where animation directly supports the explanation of medical concepts.

<sup>3</sup>CES (Complementary-Enhancement-Salience), CEE (Complementary-Enhancement-Expansion), CEPS (Complementary-Enhancement-Priority & Secondary), CNI (Complementary-Non-enhancement-Interaction), CNCom (Complementary-Non-enhancement-Combination), CNCoo (Complementary-Non-enhancement-Coordination).

For example, in the episode *What happens during a stroke?*, the narration states, “Ischemic stroke occurs When a clot blocks a vessel and brings blood flow to a halt.” Simultaneously, the screen displays a brain cross-section with a red, anthropomorphic blood clot labeled “BLOOD CLOT,” which scowls angrily while physically obstructing the vessel. This vivid and exaggerated depiction not only mirrors the narration but also amplifies its impact by portraying the clot as a disruptive force with a hostile facial expression, emphasizing the severity of the blockage. The animation enhances the spoken content and makes the pathophysiological mechanism more memorable and easier to understand, particularly for lay audiences with limited medical background.

Complementary-Enhancement-Expansion (CEE, 14.23%) and Complementary-Enhancement-Priority & Secondary (CEPS, 7.12%) have lower proportions, suggesting that expansion and primary-secondary complementary relationships are relatively less. Animation mainly serves language rather than introducing additional information.

In the episode *How does the thyroid manage your metabolism?*, the narration states, “The pituitary’s role is to sense if hormone levels in the blood are too low or too high.” At the same time, the animation gradually presents the words “Hormones” and highlights the positions of the pituitary gland and thyroid within a human silhouette. Arrows and gauges are used to visually represent how hormone signals travel between these glands. This segment is primarily marked as a Complementary-Enhancement-Extension (CEE) relationship. The animation extends the verbal explanation by visually mapping out the hormonal feedback mechanism described in the narration. The moving arrows and highlighted glands provide a clearer picture of how hormone levels are monitored in the body, enhancing the audience’s concrete understanding of the process.

At the same time, this segment also reflects features of a Complementary-Extension-Priority & Secondary (CEPS) relationship. While the animation plays an important explanatory role, the narration remains the primary mode, guiding the viewer through the logic of the hormonal regulation system. The animation supports the verbal content without overwhelming it, reinforcing the message rather than delivering new information independently.

#### **4.3.2. Relationship-LP (Language-Picture)**

Among the language-picture relationships, Complementary-Enhancement-Salience (CES, 9.23%) appears most frequently. However, its overall proportion is significantly lower than that in language-animation relationships, suggesting that pictures in TED-Ed videos play a much less integrative role than animations. Instead, pictures tend to function as visual highlights that reinforce key verbal content.

In the episode *Does stress cause pimples?*, when the narration mentions “To release cortisol, the major stress hormone,” the animation temporarily freezes on a static frame where the word “Cortisol” appears in bold next to a simplified drawing of adrenal glands. This still frame remains there for several seconds without

any motion or transition, visually emphasizing the term “Cortisol” and anchoring it firmly in the viewer’s memory.

Complementary-Enhancement-Expansion (CEE, 5.34%) and Complementary-Enhancement-Priority & Secondary (CEPS, 1.55%) occur less frequently in language-picture relationships, suggesting that pictures play a more limited role in expanding content or forming layered communicative structures. However, these two modes offer unique advantages in particular knowledge contexts. CEE relationships are particularly effective when dealing with specialized or unfamiliar concepts; by supplementing language with additional visual elements, they provide viewers with a more comprehensive understanding. CEPS relationships, meanwhile, help establish a communicative hierarchy, where language takes the lead and static images play a supporting role—guiding the viewer’s focus and aiding interpretation without distraction.

The episode *Why do people have seasonal allergies?* exemplifies both CEE and CEPS. While the narration introduces allergens, the screen presents a paused image featuring various allergen illustrations, including pollen, dust mites, and pet dander. This frame remains frozen with no moving elements, serving as a period of visual supplementation to help viewers intuitively recognize common allergens. Although both CEE and CEPS relationships can be identified here, the scene primarily functions as a Complementary-Enhancement-Expansion instance, as the image provides new semantic content not directly mentioned in the narration.

#### **4.3.3. Relationship-AP (Animation-Picture)**

Complementary-Enhancement-Salience (2.64%) is the only complementary relationship type within complementary AP. Although this proportion may seem modest, it highlights the strategic collaboration between these two visual modalities. Static images play a crucial role in reinforcing and grounding animated content, enhancing viewers’ ability to understand and retain complex biomedical concepts.

For example, in the episode *How to Grow a Bone?*, moving pictures show the cutting process, while on the right side there is a still picture of different bone shapes. This arrangement helps people see how bones grow step by step in the animation and how finished bones look in the picture. The marked parts like epiphyseal plates and cortical bone help to link the animation’s biological ideas with real body parts that we can recognize. The picture is not only just supporting the animation but also plays an important role in making the process clear by offering concrete visual details.

In the episode *Your Body vs. Implants*, the same design can be seen. When the voice says, “And it’s not just glucose monitors and insulin pumps that have this problem,” the animation shows different implants working with body tissues. At the same time, a still X-ray picture is there, showing how implants are put inside the bone structure. This helps people understand the information better by connecting the moving picture of the body’s reaction to implants with real hospital images. The X-ray also gives clear view that helps explain the complex knowledge

about body implants.

The statistical analysis further shows that other interaction modes like CEE, CEPS, or non-enhancement types such as CNI, CNCom, CNCoo are not common, which clearly tells us that TED-Ed videos are designed with purpose. Here, static images work as good visual supports, helping animation to convey ideas in a vivid way. In the meantime, the videos make sure that the main biomedical concepts stay clear without confusing viewers with too many complex relationships among different modes.

#### 4.4. Detailed Analysis of Non-Complementary Relationships

The proportion of non-complementary relationships in TED-Ed medical science videos is generally low, indicating that the videos predominantly aim for multi-modal coordination to enhance information delivery. However, in certain cases, some modes may exhibit irrelevance, missing information, or partial inclusion, which could be attributed to production strategies, visual style, or instructional design. We extract percentage data for the seven categories NOR, NOE, NON, NIWP, NIAC, NCII, and NCID<sup>4</sup> from the annotated materials and calculate their average proportions in Relationship-LA, Relationship-LP, and Relationship-AP. The processed results are presented below in **Table 5**, followed by specific examples for further analysis.

**Table 5.** The average distribution of non-complementary relationships.

Category	Relationship-LA/%	Relationship-LP/%	Relationship-AP/%
NOR	0.757	2.368	0
NOE	0	0	0
NON	0	0	0
NIWP	0	0	13.151
NIAC	0.539	0	0
NCII	0	0	0
NCID	0	0	0

##### 4.4.1. Relationship-LA (Language-Animation)

Non-complementary-Overlap-Redundancy under LA (0.757% in **Table 5**) indicates that in rare cases, language and animation may not be directly related. The animation does not align with the ongoing narration but appears on screen without directly supporting the spoken content.

In the episode *What would happen if you didn't sleep?*, the narration states, <sup>4</sup>NOR (Non-complementary-Overlap-Redundancy), NOE (Non-complementary-Overlap-Exclusion), NON (Non-complementary-Overlap-Neutralization), NIWP (Non-complementary-Inclusion-Whole & Part), NIAC (Non-complementary-Inclusion-Abstract & Concrete), NCII (Non-complementary-Contextual Interaction-Independent), and NCID (Non-complementary-Contextual Interaction-Dependent).

“So, how does our brain prevent deposits overload from happening while we sleep.” Meanwhile, the animation displays a diagram of the body system, highlighting “glymphatic system” deposits with text labels. It clearly explains glymphatic system as a cleaning system that clears out the deposits. In this case, the animation reinforces and visualizes key elements of the narration without introducing additional information, serving to strengthen comprehension rather than expand content.

Non-complementary-Inclusion-Abstract or Concrete in LA (0.539% in **Table 5**) suggests that the animation may partially align with the concept expressed in language but does not fully match it. The animation may provide additional information, or its degree of informational alignment is relatively low.

In the episode *How do the lungs do?*, the narration states, “Instead of being hollow lungs are actually spongy inside.” Meanwhile, the animation transitions from a magnified view of alveolar structures to an overhead view of a sponge, forming a visual analogy. The narration already conveys the fact that the lungs’ internal structure is like a sponge which is filled with bubbles, and the animation only makes this comparison more intuitive through a visual transformation. Even without watching the animation, listeners can still understand the concept. The animation merely makes the analogy more direct but does not provide new knowledge or data.

#### 4.4.2. Relationship-LP (Language-Picture)

Non-complementary-Overlap-Redundancy in LP (2.368% in **Table 5**) indicates that in a small number of cases, pictures do not directly correspond to language content and may serve decorative purposes rather than supporting the spoken narration.

In the episode *What would happen if you didn’t sleep?*, the narration states, “The brain’s cleaning system works by using the cerebrospinal fluid to flush away toxic byproducts.” Meanwhile, a static picture illustrating waste clearance in the brain appears on the screen, showing how cerebrospinal fluid flushes out metabolic waste. This image functions similarly to a supplementary slide in a presentation, emphasizing and reinforcing the concept rather than providing additional explanation.

#### 4.4.3. Relationship-AP (Animation-Picture)

Non-complementary-Inclusion-Whole & Part in AP (13.151% in **Table 5**) indicates that the animation and picture primarily exhibit a whole-part relationship, where the picture is often part of the animation’s content but does not fully match or directly supplement the animation’s information. This relationship suggests that the image is embedded as a part of the animation but does not enhance or reinforce the information conveyed by the animation. Instead, it reflects a parallel distribution of information, jointly constructing a complete medical concept.

In the episode *What would happen if you didn’t sleep?* the animation initially displays a full-body depiction of a human figure, then zooms in on the brain, fol-

lowed by the appearance of a close-up static image of the cerebral cortex to illustrate how sleep deprivation affects the brain. The narration explains the impact of sleep deprivation on brain function, while the animation shows the entire body, and the static brain image zooms in on a key detail. This indicates that within the Animation-Picture relationship, the animation can provide a broad contextual backdrop, while static pictures zoom in on critical details, making the information more visually clear. However, rather than being complementary, they form a hierarchical progression in information presentation.

#### 4.5. Implications for Contemporary Science Popularization Videos

This research analyzes multimodal discourse in TED-Ed medical videos in great detail, illustrating how animation, language and static pictures work together to bring a better science communication effect. The results tell us that animation usually acts as main mode to express ideas, with language and images helping and emphasizing the visual content. Also, though non-complementary relations do not happen so often, they are very important for making smooth changes, improving coherence and broadening the space for people to comprehend scientific information.

The research can bring some valuable inspiration for the production of science popularization videos:

- 1) Optimize multimodal synergy: effectively combine animation, language, and pictures to ensure complementary and enhancing relationships among modalities, thereby improving information delivery.
- 2) Leverage non-complementary relationships: strategically incorporate non-complementary elements as transitions or cognitive buffers to aid audience comprehension of complex scientific concepts.
- 3) Enhance narrative strategies: employ diverse narrative perspectives (e.g., shifts between first-person and third-person) to increase audience engagement and resonance.
- 4) Emphasize visual appeal: recognizing the growing audience expectations for visual content, create more engaging and interactive animations to enhance the viewing experience.

### 5. Conclusion

This study employed ELAN software to conduct a detailed annotation and analysis of 12 TED-Ed medical science videos, revealing the interactive relationships and information construction strategies among language expression, animation, and static images. The key findings of this study are as follows:

The interaction between animation and language, as well as between animation and pictures, generally exhibits a complementary-enhancement relationship. Among these, salience is particularly prominent. Animation often serves as the primary expressive modality, while language or pictures function as supplementary elements to reinforce and emphasize the visual content.

Complementary-non-enhancement and non-complementary relationships occur less frequently in all the videos. When present, they primarily serve as transitional mechanisms, where language or images precede, and the other modality follows, enhancing coherence and comprehension. Although these relationships may do not directly amplify each other's meaning, they are important in facilitating smooth transitions and maintaining logical consistency between different modalities.

A concise core statement is often introduced at the beginning of the video to serve as a thematic thread throughout. Although these statements are not immediately explained through narration, they are later elaborated upon through language and animation, thus forming a cohesive knowledge loop. This approach fosters self-directed thinking and learning, enhancing the interactivity and appeal of the videos.

While this study focuses on TED-Ed medical science videos, the insights are transferable to other forms of educational content, such as environmental science, psychology, or biology. The core principle of modality complementarity—where animation visually supports narration—can enhance learning across domains. Especially in STEM education, where abstract concepts often challenge comprehension, multimodal coordination serves as a vital tool for improving engagement and retention.

Compared to science podcasts or infographics, TED-Ed's multimodal video format offers richer sensory engagement. While podcasts rely solely on verbal explanation, limiting the visualization of biological phenomena, infographics may present static information without the narrative coherence of a video. TED-Ed's approach—animated storytelling with tightly scripted narration—bridges this gap by offering both cognitive depth and emotional appeal.

In conclusion, this research can provide theoretical support and practical guidance for the multimodal design of science popularization videos, facilitating effective dissemination of scientific knowledge and the enhancement of public scientific literacy.

## **Acknowledgements**

The authors gratefully acknowledge the research project “Internet-based Collaborative Learning Hub—Home for Personalized ESP Lifelong Learning” supported by China's National Program of Undergraduate Innovation and Entrepreneurship (No. 202313022022).

## **Conflicts of Interest**

The authors declare no conflicts of interest regarding the publication of this paper.

## **References**

Alhassan, R. K., Abdul-Fatawu, A., Adzimah-Yeboah, B., Nyaledzigbor, W., Agana, S., & Mwini-Nyaledzigbor, P. P. (2019). Determinants of Use of Mobile Phones for Sexually

- Transmitted Infections (STIs) Education and Prevention among Adolescents and Young Adult Population in Ghana: Implications of Public Health Policy and Interventions Design. *Reproductive Health*, 16, Article No. 120.  
<https://doi.org/10.1186/s12978-019-0763-0>
- Forceville, C. (1996). *Pictorial Metaphor in Advertising*. Routledge.
- Forceville, C. (2009). The Role of Non-Verbal Sound and Music in Multimodal Metaphor. In *Multimodal Metaphor* (pp. 383-402). Mouton de Gruyter.  
<https://doi.org/10.1515/9783110215366.6.383>
- Kress, G., & Leeuwen, T. (1996). *Reading Images: The Grammar of Visual Design*. Routledge.
- Li, Z. Z. (2003). Social Semiotic Approach to Multimodal Discourse. *Foreign Languages Research*, 5, 1-8.
- Liu, H., & Zhao, C. G. (2020). Health Science Popularization via Short-Video and Live Video Streaming Platforms. *Chinese Preventive Medicine*, 21, 1298-1301.
- Liu, J., & Hu, K. B. (2015). The Compilation and Use of Multimodal Interpreting Corpora. *Foreign Languages in China*, 12, 77-85.
- Norris, S. (2004). *Analyzing Multimodal Interaction: A Methodological Framework*. Routledge.
- Romano, A., Fiori, F., Petrucci, M., Della Vella, F., & Serpico, R. (2022). YouTube Content Analysis as a Means of Information in Oral Medicine: A Systematic Review of the Literature. *International Journal of Environmental Research and Public Health*, 19, Article 5451. <https://doi.org/10.3390/ijerph19095451>
- Tereszkiewicz, A. (2023). Engagement Strategies on Medical YouTube Channels. *Studia Linguistica Universitatis Jagellonicae Cracoviensis*, 140, 139-164.  
<https://doi.org/10.4467/20834624sl.23.007.17756>
- Wang, Y. G., & Fan, Q. L. (2019). Problems in the Health Dissemination of the Short Video Platform of Tik Tok and Improvement Paths. *Journal of Chang'an University (Social Science Edition)*, 21, 53-60.
- Xin, Z. Y. (2008). A New Developments in Discourse Analysis: Multimodal Discourse Analysis. *Social Science Journal*, 5, 208-211.
- Zhang, D. L. (2009). On a Synthetic Theoretical Framework for Multimodal Discourse Analysis. *Foreign Languages in China*, 6, 24-30.
- Zhang, D. L., & Zhang, K. (2022). Developing an Integrated Framework for Multimodal Critical (Positive) Discourse Analysis. *Foreign Language Education*, 43, 1-8.
- Zhu, Y. S. (2007). Theory and Methodology of Multimodal Discourse Analysis. *Foreign Language Research*, 5, 82-86.