



The Multimodal Construction of Affective Stance in ASMR Videos

Qian Wang

College of Humanities, Tiangong University, Tianjin, China

Email: 448018382@qq.com

How to cite this paper: Wang, Q. (2022) The Multimodal Construction of Affective Stance in ASMR Videos. *Open Access Library Journal*, 9: e9070.
<https://doi.org/10.4236/oalib.1109070>

Received: July 1, 2022

Accepted: July 24, 2022

Published: July 27, 2022

Copyright © 2022 by author(s) and Open Access Library 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

Multimedia technology develops and brings considerable changes in interpersonal communication. Nowadays, communicating with non-verbal languages such as pictures, texts, music, etc. is a common occurrence. And scholars have started to construe the meaning behind these new discourses which are carried by more than one mode. As a result, it unseals the gate to the analysis of multimodal discourses. ASMR videos, autonomic sensory meridian response videos related to aural and visual stimuli, are born to be multimodal discourses, but are rarely studied. Therefore, 5 ASMR videos on YouTube are chosen as the data for this thesis. And backed up by the foundation stones—visual grammar and audio grammar, an analysis of the selected videos was made in an audio-visual mode in order to elaborate on the transmittal of affective stance in them. The present study shows that ASMR performers all index intimate affective stances. Besides, performers have different focuses in their video making. Some focus on visual stimuli, such as on putting contact lenses, while others focus on aural stimuli with either high pitch or broad pitch range to stimulate the audience or low pitch and narrow pitch range to create a relaxing atmosphere. By analyzing the selected videos, the present study may serve as an expansion of the construction of affective stance and an exploration of ASMR in linguistics. Meanwhile, it provides suggestions for future studies in ASMR.

Subject Areas

Linguistics, Multimodal Discourse Analysis

Keywords

Multimodal Discourse Analysis, Affective Stance, ASMR Videos, Visual Grammar, Audio Grammar, YouTube, Interactive Meaning

1. Introduction

1.1. ASMR

ASMR, autonomic sensory meridian response, is usually described as a tingling or shivering experience on the scalp or along the spine. And it can be felt when exposed to whispering, tapping, and other visual and aural stimuli (Barratt & Davis, 2015 [1]; Michele, 2021 [2]). While this sensation is not new, ASMR has only recently been recognized and named (Starr, R. L., Wang, T. X. & Go, C. 2020 [3]). And the term ASMR is coined in 2010 (Allen, 2016) [4]. As a result, ASMR videos pour in and they mainly use situational simulation or aural stimuli to get people immersed in a virtual environment.

Research on them have also begun to sprout accordingly. The author searched ASMR in ScienceDirect and CNKI respectively. 1286 results were presented in ScienceDirect. Of which, 743 are about Medicine and Dentistry, 317 are about Pharmacology, Toxicology and Pharmaceutical Science. However, in linguistics, ASMR videos are seldom studied.

As for CNKI, the author obtained only 18 results. And there is only 1 research in Foreign Languages and Literatures. It can be roughly concluded that ASMR researches in linguistics are in need. And stance may be a good starting point.

1.2. Affective Stance

Stance or stance taking, according to Ronald Wardhaugh and Janet M. Fuller (2015) [5], is “*the use of language to position oneself with regard to other interlocutors as well as attitudes and ideologies being discussed*”. Just as Bucholtz and Hall proposed in 2005 [6], the stance is an important language resource to indicate social identity. And it can be further divided into epistemic stance and affective stance (Jaffe, 2009) [7]. The epistemic stance focuses on the relationship between the speaker and what is said while the affective stance is about the relationship between speaker and receiver (Kiesling, 2009: 172) [8] which this paper focuses on.

There has been a lot of research on epistemic stance in the field of discourse analysis (Heritage, 2012) [9]. Affective stance, however, is rarely studied (Xia Jie & Chen Xinren, 2021) [10]. Just as Damasio (2018) [11] mentioned, affections are omnipresent in life, but they have been neglected in scientific research for a long time. An individual’s way of speaking is a performative act of participating in the construction of social identity (Zhao Peng & Tian Hailong, 2022) [12]. And according to D’Onofrio & Eckert (2021) [13], affection, previously considered as individual differences, is now also studied as social variable. Besides, the author also searched them in ScienceDirect and CNKI respectively, the results still match the former studies. Therefore, further study should be made on the affective stance (Alba-Juez & Mackenzie, 2019) [14].

In short, more researches on affective stance are in need. And ASMR, born with visual mode and audio mode, is a new thing to be reckoned with. Therefore, this paper may serve as an exploration of multimodal discourse analysis about ASMR, and be propitious to explain the transmittal of affective stance in

ASMR.

As the aim of this thesis is to unveil the transmittal of affective stance in ASMR videos, the study addresses the following research questions:

- 1) What kind of affective stance is constructed in ASMR?
- 2) How is affective stance constructed by audio-visual mode in ASMR?

2. Methodology

2.1. Approach

Based on visual grammar (2006) and audio grammar (1999), the study will employ both qualitative and quantitative approaches to hold a case study of the affective stance in 5 ASMR videos.

2.2. Data Collection

The data was collected from YouTube according to the score. For the sake of concreteness, the author chose the former 5 videos whose length is all less than 4 minutes to make an analysis (Table 1).

3. Theoretical Framework

3.1. Interactive Meaning in Visual Grammar

With the support of three functions in Halliday's functional grammar (2014) [15], Kress and Van Leeuwen initiated three meanings accordingly to deal with visual mode. To analyze the affective stance, this thesis will focus on interactive meaning because of its communicative feature.

This paper focuses on the relationship between the represented symbols and audience, and relates to contact, social distance, perspective (Kress & van Leeuwen, 2006) [16].

3.1.1. Contact

In visual grammar, visual discourses express their meaning through image act which consists of demand act and offer act.

Demand act is typically featured with the direct gaze from the represented symbols to audience. In this situation, the represented symbols can be persons, animals, or any personified objects. They gaze and talk with audience to demand

Table 1. Data samples.

No.	Title	Length
1	ASMR Spicy Food _ Chinese food and cooking, dinner, drink, BBQ restaurant	14"
2	60 triggers in 1 seconds asmr	1'00"
3	Asmr haircut in 1 minute	59"
4	[ASMR] 100+ TRIGGERS!	1'54"
5	ASMR Put contact lenses in your eyes! 1min _FASTEST EVER_	51"

audience' attention and assume the presence of audience which invites audience to be involved in imaginary relations. And gestures or facial expressions are often used to amplify demand act.

One the contrary, offer act is featured without direct visual contact. The represented symbols can be any creatures or objects and audience are just bystanders or passers-by to see what is happening.

In simple words, demand act refers to that represented symbols demand attention from audience while offer act means the represented symbols offer information to audience.

3.1.2. Social Distance

Social distance can reveal the level of affinity between participants. Like in daily life, shorter distance usually means closer relation. And in visual discourses, the choices of close-up shot, medium shot and long shot suggest the interpersonal relations from intimacy to separation.

Close-up shot, focusing on heads and shoulders or participants' bodies above waists, suggests intimate or personal relation. Medium shot, showing the whole bodies of participants, signifies social distance when people conduct business or other interactive activities. Long shot, exhibiting the whole bodies and background, identifies the distance between strangers.

3.1.3. Perspective

Perspective, related to the angle, can be mainly split into horizontal angle and vertical angle. And horizontal angle covers frontal angle and oblique angle. When the represented symbols face to audience directly, a front angle is constructed and it seems that audience are involved in the same situation with the represented symbols. On the contrary, oblique angle signifies separation.

As for vertical angle, it covers high angle, eye-level angle, and low angle. They can imply the power relations. With a high angle, audience needs to look down to see the represented symbols, which implies the represented symbols don't have more power than the audience. An eye-level angle suggests an equality. And a low angle forces the audience to look up to see the represented symbols which signifies they are more powerful than the audience.

3.2. Interactive Meaning in Audio Grammar

In multimodal discourses, images and texts are mainly analyzed by numerous scholars. Sounds, however, are rarely studied for the difficulty of transcription. As for related theories, van Leeuwen has proposed the audio grammar in his work *Speech, Music, Sound* (1999) [17] based on the visual grammar. He explores the similarities of audio resources such as music, speeches, videos and other sounds.

Rules of perspective, social distance, and melody in audio grammar are chosen to analyze the 5 ASMR videos.

3.2.1. Perspective

Perspective in audio resources can be split into figure, ground, and field, which

suggest the interpersonal relations from intimacy to separation. And it can be realized by the degree of loudness of the sounds that happen simultaneously, regardless of their producers are the people or objects.

Figure refers to the most prominent sound, and receivers of it always identify and react to them immediately in their social world. Ground is the less prominent sound we take for granted in the listener's social world, that is to say, it resembles a background which is only noticed when it disappears. As for field, if a sound is placed as field, it is only treated as an existing sound in the receivers' physical world rather than social world.

3.2.2. Social Distance

As for social distance, it has 5 types as it in speeches. In speeches, social distance can be classified to intimate distance, personal distance, informal distance, formal distance and public distance. And as distance grows, the sound always gets louder, higher and sharper.

In audio grammar, intimate distance is always related to whispering or maximally soft sounds. Personal distance, similarly, is pronounced by a low pitch and volume of a tend and relaxed sound louder than whispering. As for informal distance, it is symbolized by a medium sound louder than personal distance. For formal distance, it is usually created by a louder, higher and sharper sound in a formal occasion. Finally, public distance is often recognized by the maximally loud sounds.

3.2.3. Melody

Melody, the wave pattern of sounds consist of pitch movement, pitch range, and pitch level, is also noticed to have a noticeable impact on the construction of interactive meaning (Van Leeuwen, 1999) [17].

Pitch movement can be marked as falling or rising. When the pitch has more rises, the participants including both speakers and receivers are always more active, while the more the pitch falls, the quieter the participants will be, and they may get immersed in relaxation and meditation. Pitch range can infer the emotion of the speaker and influence the receiver's emotion in turn. For instance, broad pitch range is often correlated with a big emotive undulation such as incitement or astonishment while narrow pitch range mainly expresses more restrained emotions like misery and calm. Finally, pitch level refers to the degree of highness in a sound and a higher voice is usually associated with higher power while a soft one can show the intimacy between the speaker and receiver.

4. Data Analysis

4.1. Affective Stance in Visual Mode

The author drew on the Elan 6.3 to mark and count the proportion of the concrete presentation strategy. According to visual grammar, contact consists of demand act and offer act. And all the gaze parts of the performer are counted as demand act while other parts are offer act. As for social distance, it can be di-

vided into close-up shot, medium shot, and long shot. Among them, what focus on participants' bodies above waists are counted as close-up shots, while medium shots show the whole bodies of participants, and long shots exhibit the whole bodies and background. Finally, perspective can be mainly split into horizontal angle and vertical angle. And horizontal angle covers frontal angle and oblique angle. When the represented symbols face to audience directly, a front angle is constructed. On the contrary, oblique angle signifies separation. As for vertical angle, it covers high angle, eye-level angle, and low angle. With a high angle, audience needs to look down to see the represented symbols. An eye-level angle suggests an equality. And a low angle forces the audience to look up to see the represented symbols.

Therefore, the results of Elan 6.3 are shown as the following **Table 2**.

As we mentioned in Section 3.1.1, demand act means the direct gaze. In the former 4 videos, gaze is seldom or even never employed in order to create a relaxing atmosphere. The last video, however, is always gazing the audience to simulate the situation that the performer is putting contact lenses in the audience's eyes in order to get them involved the imaginary situation as shown in **Figure 1**.

As for social distance and perspective, all the performers are sitting on a chair without walking, and the audience can only see the upper part of their body frontally as shown in **Figure 1** and **Figure 2**, which seems that the audience is

Table 2. Proportions in 5 videos.

No.	Contact		Social distance			Perspective				
	Offer act	Demand act	Close-up shot	Medium shot	Long shot	Frontal angle	Oblique angle	High angle	Eye-level angle	Low angle
1	81.12%	18.88%	100%	0	0	100%	0	0	100%	0
2	100%	0	100%	0	0	100%	0	0	100%	0
3	94.40%	5.60%	100%	0	0	100%	0	0	100%	0
4	100%	0	100%	0	0	100%	0	0	100%	0
5	4.54%	95.46%	100%	0	0	100%	0	0	100%	0



Figure 1. Gaze and close-up shot.

involved in the same situation and suggests an intimate relation between the performer and audience. Besides, the eye-level angle all makes up 100% in these videos to show an equal status between the performer and audience and shorten the distance between them.

4.2. Affective Stance in Audio Mode

In audio mode, the sounds are all analyzed in general. According to audio grammar, perspective in audio resources can be split into figure, ground, and field. Figure refers to the most prominent sound, and receivers of it always identify and react to them immediately in their social world. Ground is the less prominent sound we take for granted in the listener's social world. As for field, it is only treated as an existing sound in the receivers' physical world rather than social world. For social distance, intimate distance is always related to whispering or maximally soft sounds. Personal distance, similarly, is pronounced by a low pitch and volume of a tend and relaxed sound louder than whispering. As for informal distance, it is symbolized by a medium sound louder than personal distance. For formal distance, it is usually created by a louder, higher and sharper sound in a formal occasion. Finally, public distance is often recognized by the maximally loud sounds.

Therefore, the author made a rough classification as shown below (Table 3).

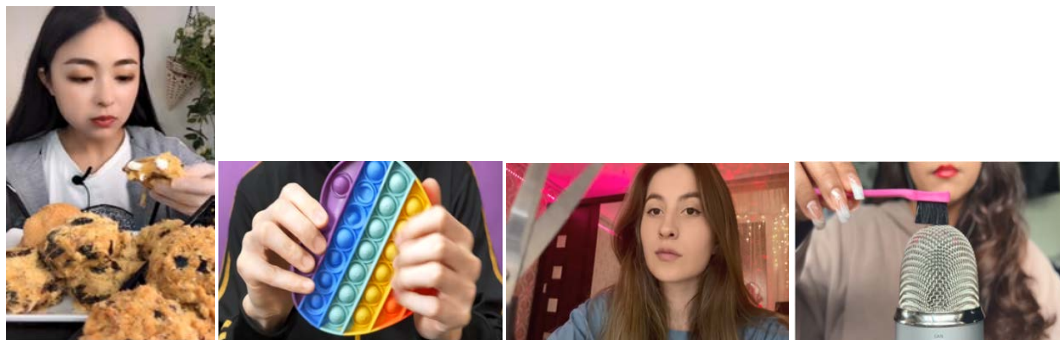


Figure 2. Close-up shots and eye-level angle.

Table 3. Classification of sounds in 5 videos.

No.	Perspective			Social distance				
	Figure	Ground	Field	Intimate	Personal	Informal	Formal	Public
1	Chewing sounds	Music	A child's screams	Chewing sounds	0	0	Music	A child's screams
2	Tapping sounds	0	0	Tapping sounds	0	0	0	0
3	Sounds of tools	Sounds of machines	0	Sounds of tools	0	0	Sounds of machines	0
4	Sounds of nails	0	0	Sounds of nails	0	0	0	0
5	Whispering & sounds of tools	0	0	Whispering & sounds of tools	0	0	0	0

In these 5 videos, 3 of them just have the figure which the performers demand attention for in order to get the listener involved in the simulated situation to show intimacy. For the first video, it is a mukbang video with music. The music sound can construct an atmosphere of relaxing while the screams of a child may be recorded by accident, because the performer pays no attention to the child. In addition, the sounds of machines in video 3 show the listener more details in the simulated situation in order to be more authentic. In brief, these videos all show an intimacy between the performer and listener.

As for melody, the author analyzed the pitch movement, pitch range and pitch level according to the graphic waves in Elan 6.3 as shown below (Figures 3-7). And melody refers to the wave pattern of sounds which consist of pitch movement, pitch range, and pitch level. Pitch movement can be marked as falling or rising. Pitch range can be divided into broad and narrow range. Finally, pitch level refers to the degree of highness in a sound.

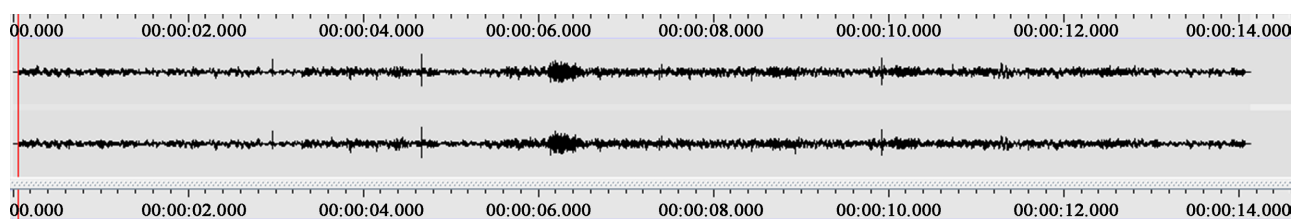


Figure 3. Wave pattern of video 1.

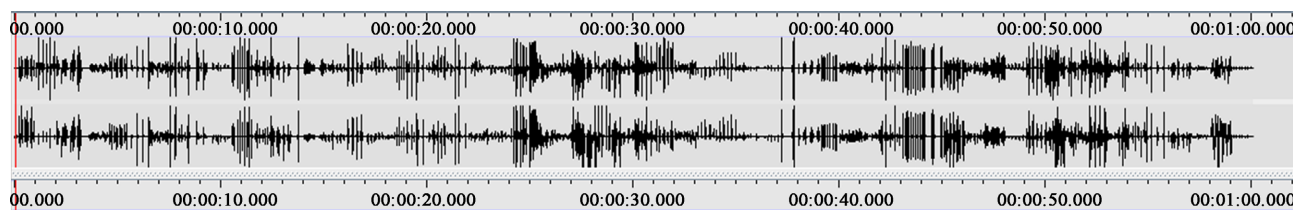


Figure 4. Wave pattern of video 2.

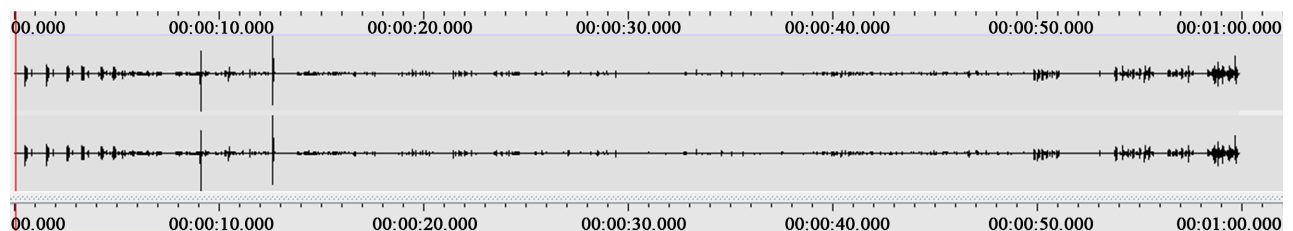


Figure 5. Wave pattern of video 3.

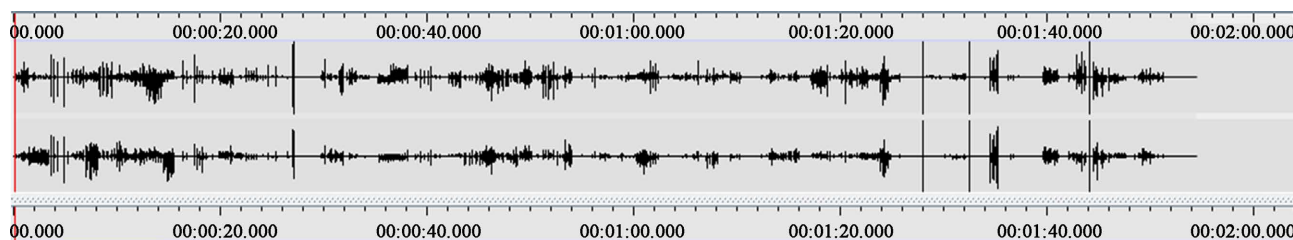


Figure 6. Wave pattern of video 4.

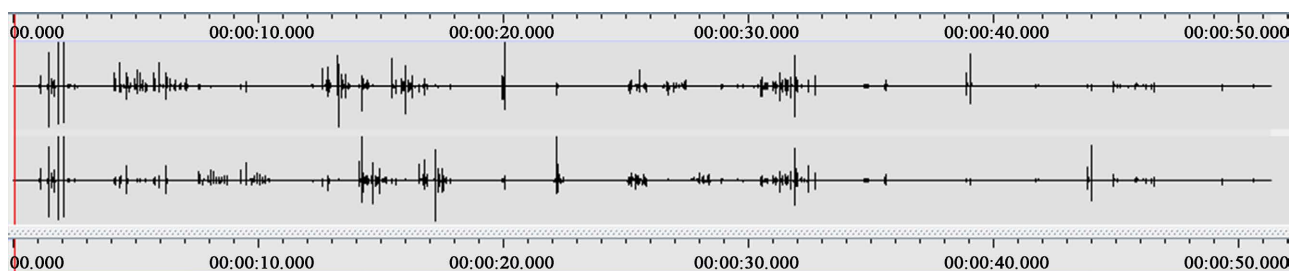


Figure 7. Wave pattern of video 5.

Table 4. Melody.

No.	Melody		
	Pitch movement	Pitch range	Pitch level
1	stable	narrow	middle
2	more rising	broad	high
3	stable	narrow	low
4	more rising	middle	middle
5	more rising	middle	middle

From their waves, the author made a summary as below (Table 4).

Videos 1 and 3 both show stable pitch movement and narrow pitch range, that's probably because they use ground sounds whose pitch is stable with both rising and falling in order to create a peaceful atmosphere. However, the ground sounds in video 1 is music, so the pitch level is middle while in video 3, there are the sounds of machines whose pitch is as low as white noise.

As for videos 2, 4, and 5, they have more rising because only a figure is used to attract the listener's attention, and what the listener can receive is only figure of silence. In video 2, the pitch level is the highest of the 5 videos, and its pitch range is the broadest. The reason for this is that the performer has tapped plenty of objects rapidly and there are few lapses of time between 2 objects. Video 4, however, is also around the sounds of a large number of objects, the performer chooses a gentler and slower way to produce the sounds and the lapses of time are longer. The listener therefore can feel more relaxing because of the middle pitch and a middle wave range of pitch.

Finally, video 5 which has the same pitch features as video 4, can also symbolize an intimate relationship between the performer and listener.

5. Conclusions

As an innovative genre that highlights the sensual experience of the audience, ASMR videos deserve the attention of sociolinguistics.

In view of limited time and space, this study makes a general analysis of affective distance in ASMR videos, and finds that although all the selected videos are found to index intimate affective stance, the performers have different focuses.

Some focus on visual stimuli, such as putting on contact lenses, while others focus on aural stimuli with either high pitch or broad pitch range to stimulate the audience or low pitch and narrow pitch range to create a relaxing atmosphere. For concreteness, researchers in the future can shine the light on one kind of them or videos of one performer.

In addition, audio grammar fits general analysis, phonetics can be utilized for more detailed analysis and the software Praat may come in handy to record the exact number of pitch levels.

Besides, this thesis pays no attention to the distinctions between the human voice and the sounds of objects. For instance, when people get closer to a microphone, they tend to lower their voice unconsciously, while objects won't. Whether the distinctions can cause different responses from the audience also deserves attention.

Conflicts of Interest

The author declares no conflicts of interest.

References

- [1] Barratt, E.L. and Davis, N.J. (2015) Autonomous Sensory Meridian Response (ASMR): A Flow-Like Mental State. *PeerJ*, **3**, e851. <https://doi.org/10.7717/peerj.851>
- [2] Michele, Z. (2021) Ambient Affiliation in Comments on YouTube Videos: Communing around Values about ASMR. *Journal of Foreign Languages*, **44**, 21-40.
- [3] Starr, R.L., Wang, T.X. and Go, C. (2020) Sexuality vs. Sensuality: The Multimodal Construction of Affective Stance in Chinese ASMR Performances. *Journal of Sociolinguistics*, **24**, 492-513. <https://doi.org/10.1111/josl.12410>
- [4] Allen, J. (2016) Interview with Jennifer Allen, the Woman Who Coined the Term, "Autonomous Sensory Meridian Response" (ASMR). ASMR University, New York. <https://asmruniversity.com/2016/05/17/jennifer-allen-interview-coined-asmr>
- [5] Ronald, W. and Janet, M.F. (2015) An Introduction to Sociolinguistics. John Wiley and Sons, Inc., New York.
- [6] Bucholtz, M. and Hall, K. (2005) Identity and Interaction: A Sociocultural Linguistic Approach. *Discourse Studies*, **7**, 585-614. <https://doi.org/10.1177/1461445605054407>
- [7] Jaffe, A. (2009) Stance: Sociolinguistic Perspective. Oxford University Press, Oxford. <https://doi.org/10.1093/acprof:oso/9780195331646.001.0001>
- [8] Kiesling, S. (2009) Style as Stance: Can Stance Be the Primary Explanation for Patterns of Sociolinguistic Variation? In: Jaffe, A., Ed., *Sociolinguistic Perspectives on Stance*, Oxford University Press, Oxford, 171-194. <https://doi.org/10.1093/acprof:oso/9780195331646.003.0008>
- [9] Heritage, J. (2012) Epistemics in Action: Action Formation and Territories of Knowledge. *Research on Language and Social Interaction*, **45**, 1-29. <https://doi.org/10.1080/08351813.2012.646684>
- [10] Xia, J. and Chen, X.R. (2021) A Citespace Based Analysis of Linguistic Stance Research Abroad. *Foreign Language and Literature Research (Series)*, **12**, 162-177.
- [11] Damasio, A. (2018) The Strange Order of Things. Pantheon Books, New York.
- [12] Zhao, P. and Tian, H.L. (2022) The New Development of Variation Sociolinguistic

-
- Research. *Modern Foreign Languages*, **45**, 137-147.
- [13] D'Onofrio, A. and Eckert, P. (2021) Affect and Iconicity in Phonological Variation. *Language in Society*, **50**, 29-51. <https://doi.org/10.1017/S0047404520000871>
- [14] Alba-Juez, L. and Mackenzie, L. (2019) Emotional Processes in Discourse. In: Alba-Juez, L. and Mackenzie, L., Eds., *Emotion in Discourse*, John Benjamins, Amsterdam/Philadelphia, 3-27. <https://doi.org/10.1075/pbns.302.01alb>
- [15] Halliday, M.A.K. (2014) Halliday's Introduction to Functional Grammar. Routledge, London/New York. <https://doi.org/10.4324/9780203783771>
- [16] Kress, G. and van Leeuwen (2006) Reading Images: The Grammar of Visual Design. Routledge, London/New York. <https://doi.org/10.4324/9780203619728>
- [17] Van Leeuwen, T. (1999) Speech, Music, Sound. Macmillan, London. <https://doi.org/10.1007/978-1-349-27700-1>