www.volsu.ru

МОДЕЛИРОВАНИЕ ПАРАМЕТРОВ ТЕКСТА ДЛЯ РАЗРАБОТКИ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

DOI: https://doi.org/10.15688/jvolsu2.2024.5.12

UDC 81'42
LBC 81.055.1

Submitted: 03.04.2024
Accepted: 20.08.2024

EXPERIMENTAL METHODS OF EXPLORING MULTIMODAL DISCOURSE: CROSSMODAL ALIGNMENT [1]

Mariya I. Kiose
Moscow State Linguistic University, Moscow, Russia;
Institute of Linguistics of the Russian Academy of Sciences, Moscow, Russia

Vadim O. Potekhin
Diplomatic Academy of the Russian Ministry of Foreign Affairs, Moscow, Russia;
Moscow State Linguistic University, Moscow, Russia

Oleg D. Zubkov
Moscow State Linguistic University, Moscow, Russia

Abstract. The study advances Crossmodal Alignment Framework to explore multimodal discourse in its three formats – semiotic, communicative and perceptive – via multimodal experiment. It considers the alignment patterns obtained from two semiotic modes (text and image), transferred in two communicative modes (speech and gesture), sensed by two perception modes (visual and audial). The common research framework determines the patterns as modulated by discourse tasks. The study features the results of multimodal experiments with the participants engaged in three discourse tasks: 1) receptive, which presumes obtaining information from text and image stimuli; semiotic alignment patterns are identified indirectly via participants' gaze response; 2) productive, in which the participants communicate the information in monological format; communicative alignment patterns are identified directly via their speech and gesture; 3) receptive-productive, which presupposes the participants perceive information visually and audially; alignment patterns are identified directly via participants' gaze behavior contingent on the stimuli areas of interest and indirectly via their speech response. Data analysis allows to determine and scale the degree of crossmodal alignment to discourse tasks, which helps identify the input of each mode to solving these tasks. The research framework and obtained results contribute to further development of multimodal discourse methods.

Key words: multimodal experiment, crossmodal alignment, discourse task, information construal, semiotic mode, communicative mode, perceptive mode, alignment patterns.

Citation. Kiose M.I., Potekhin V.O., Zubkov O.D. Experimental Methods of Exploring Multimodal Discourse: Crossmodal Alignment. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Yazykoznanie* [Science Journal of Volgograd State University. Linguistics], 2024, vol. 23, no. 5, pp. 149-160. DOI: https://doi.org/10.15688/jvolsu2.2024.5.12

УДК 81’42
ББК 81.055.1

Дата поступления статьи: 03.04.2024
Дата принятия статьи: 20.08.2024

ЭКСПЕРИМЕНТАЛЬНЫЕ МЕТОДЫ ИССЛЕДОВАНИЯ МУЛЬТИМОДАЛЬНОГО ДИСКУРСА: КРОСС-МОДАЛЬНАЯ АДАПТАЦИЯ [1]

Мария Ивановна Киосе
Московский государственный лингвистический университет, г. Москва, Россия;
Институт языкознания РАН, г. Москва, Россия

Вадим Олегович Потехин
Дипломатическая академия МИД России, г. Москва, Россия;
Московский государственный лингвистический университет, г. Москва, Россия

Олег Дмитриевич Зубков
Московский государственный лингвистический университет, г. Москва, Россия

Аннотация. Исследование нацелено на разработку методологии кросс-модальной адаптации для экспериментального изучения мультимодального дискурса в трех форматах его реализации: семиотическом, коммуникативном и перцептивном. Изучению подвергаются особенности конструирования дискурсивной информации, получаемой из двух семиотических модальностей (текста и изображения), транслируемой двумя коммуникативными модальностями (речи и жеста), воспринимаемой двумя перцептивными модальностями (визуальной и слуховой). Общее методологическое решение для их анализа заключается в установлении моделей адаптации этих модальностей к дискурсивным задачам. Материалом являются результаты трех экспериментов, в ходе которых участники решали три дискурсивные задачи: 1) рецептивную – извлечение информации из текста и изображения; оценка адаптации осуществлялась через отклик на нее в глазодвигательном поведении; 2) продуктивную – передачу информации в монологической коммуникации; оценка адаптации осуществлялась с помощью речи и жеста; 3) одновременно рецептивную и продуктивную – передачу информации, извлекаемой визуально и на слух; оценка адаптации осуществлялась в глазодвигательном поведении, сопряженном с зонами интереса стимула, и в речи – через коммуникативную модальность, смежную со слуховой перцептивной модальностью. Анализ полученных данных позволяет установить и ранжировать степень кросс-модальной адаптации к дискурсивным задачам, что в свою очередь определяет роль каждой модальности в их решении. Полученные результаты вносят вклад в развитие мультимодальной дискурсологии. Методология кросс-модальной адаптации разработана М.И. Киосе; эмпирические результаты получены и обработаны М.И. Киосе (рецептивная задача), В.О. Потехиным (продуктивная задача), О.Д. Зубковым (рецептивно-продуктивная задача).

Ключевые слова: мультимодальный эксперимент, кросс-модальная адаптация, дискурсивная задача, конструирование информации, семиотическая модальность, коммуникативная модальность, перцептивная модальность, паттерны согласования.

Цитирование. Киосе М. И., Потехин В. О., Зубков О. Д. Экспериментальные методы исследования мультимодального дискурса: кросс-модальная адаптация // Вестник Волгоградского государственного университета. Серия 2, Языкознание. – 2024. – Т. 23, № 5. – С. 149–160. – (На англ. яз.). – DOI: https://doi.org/10.15688/jvolsu2.2024.5.12

Introduction

Multimodal studies offer insights into exploring different types of modes. The first direction developed mostly in Systemic Functional Linguistics addresses the use of semiotic modes analyzing both verbal and nonverbal means of communicating information [van Leeuwen, 2022]. The second direction relates to the application of communicative modes like speech and gesture [Cienki, Iriskhanova, 2020; Iriskhanova, 2021] appealing to McNeill’s theory of growth points [McNeill, 2006] which reveal the cognitive nature of interaction in communicating information. The third direction, which is less considered, explores the alignment of perceptual modes, mostly aural and visual in information intake [Divjak, Milin, Medimorec, 2020]. These three directions are commonly

developed independently due to different nature of modes. Meanwhile, experimental studies of multimodal discourse face an urgent need to consider the interrelation of different mode types, i.e. crossmodality. Until recently, the methods to explore the crossmodal alignment have been developed to solve single research tasks [Keller et al., 2023], which means that the procedure applied in one research case cannot be adopted to deal with another case. We presume that advancing a common research framework to explore the alignment of different mode types can solve the problem on a systemic basis.

The assumption which underlies Crossmodal Alignment Framework is that mode alignment is dependent on the discourse tasks which conceptionally and pragmatically motivate it and which consequently may serve as the growth points [McNeill, 2006] driving information construal. Their discourse markers in speech and gesture can be directly observed in communicative behavior which allows to determine the alignment patterns. Meanwhile, the alignment of semiotic and perceptual modes or of perceptual, semiotic and communicative modes cannot be observed in this way. In [Divjak, Milin, Medimorec, 2020], for instance, the gaze behavior of experiment participants was observed contingent on the three areas of interest on the stimulus, while the participants were subjected to aurally presented information. In this case, the alignment was attested by using the gaze behavior metrics (rather, its changes), mediated by event construal in aurally presented language and in the visually presented stimulus, which means that the alignment of visual perceptual mode with aural perceptual mode was attested indirectly via communicative mode (speech). Similarly, in [Kiose et al., 2023] the alignment of semiotic modes, text and image was explored via the foregrounding features in event construal which mediate the gaze behavior of the viewers, thus in this case semiotic alignment was attested indirectly via perceptual mode. In [Chernigovskaya, Petrova (eds.), 2018] the results of multiple experimental studies are presented, where the participants' gaze behavior served to identify the alignment of text and image grounded on specific language features and specific features of image like colour and shapes.

These observations allow to claim that crossmodal studies presume that information construal is multi-directional, representing information and perceiving it, or representing, perceiving and communicating (transferring) it, which means that in case information construal is bidirectional or multidirectional, we can identify the crossmodal alignment patterns only indirectly, via a concomitant mode. Consequently, alignment identification can be explored in different discourse tasks aimed at receiving/obtaining information and producing/communicating/transferring it. This observation serves to create the common research framework for crossmodal experimental studies. In Table 1 we present the discourse task dependent on Crossmodal Alignment Framework which allows to explore both direct and indirect alignment patterns in experiments.

In this study, we address three types of multimodality in the experiments where the use of modes is stimulated by a discourse task – receptive, productive and receptive-productive. We expect that an integral view of multimodal experimental research developed in Crossmodal Alignment Framework will allow to determine and scale the multimodal alignment patterns in semiotic, communicative and perception modes.

Table 1. **Crossmodal Alignment Framework**

Multimodality type	Identification regime	Discourse task	Mode types	Examples of modes
Semiotic multimodality	Indirectly	Receptive	Via perception mode	Gaze behavior
		Productive	Via communicative mode	Speech, gesture, dance, drawing
Communicative multimodality	Directly	Productive	Via communicative mode	Speech, gesture, dance, drawing
	Indirectly	Receptive-productive	Via semiotic mode	Text, stative image, dynamic scene
		Receptive	Via perception mode	Gaze behavior
Perceptual multimodality	Directly	Receptive	Via perception mode	Gaze behavior
	Indirectly	Receptive-productive	Via semiotic mode	Text, stative image, dynamic scene
		Productive	Via communicative mode	Speech, gesture, dance, drawing

Materials and methods

To verify the Crossmodal Alignment Framework, we use the data obtained in three experimental studies testing semiotic, communicative and perception multimodality in receptive, productive and receptive-productive discourse tasks.

Experiment 1. Receptive discourse task

In Experiment 1 we determined the patterns of semiotic alignment in text and image areas of interest via the gaze behavior. Fifteen students were instructed to examine five stimuli on the computer screen (each stimulus for 15 sec, SMI red-x eye tracker was applied) and then to present a detailed account on the extracted information during the next 30 sec (blank page was demonstrated on the screen); therefore, they were subjected to a receptive task followed by a productive task (not related here). We presumed that in a receptive task the participants might be attracted to most foregrounded features of the stimuli, therefore, we addressed the typologies of foregrounding cues [Iriskhanova, 2014; Kiose et al., 2023]: 1) verbal mode cues of Graphic and orthographic foregrounding and Linguistic foregrounding, 2) pictorial mode cues of Foregrounding in image technique and Foregrounding in image colour.

Figure 1 manifests a fragment of stimulus 6[2] which was segmented into text areas of interest AOI 001 – AOI 007 and image areas of interest AOI 014 – AOI 018.

In terms of Graphic and orthographic foregrounding in text, all text AOIs are in capital font, in AOI 3 the word stress is additionally introduced, AOIs 1–3 contain non-standart orphography in *Тимофеича*, *Мокеича*, *Патрикеича* (intead of *Тимофеевича*, *Мокеевича*, *Патрикеевича*), AOIs 1 and 5 start with an interval, all AOIs apart from AOI 3 contain punctuation marks. In terms of Linguistic foregrounding, AOIs 2, 3, 6, 7 are phonetically foregrounded containing rhyming words. Morphological and lexical foregrounding is observed in several cases. AOIs 1–3 contain proper names, AOIs 1–3 manifest occasional word formation (which complies with non-standart orphography), AOIs 1, 2, 3, 5, 7 manifest different types of code-shifting (shifting to professional military terminology in *мобилизация*, *стан*, shifting to personalization by using proper names for animals). AOIs 4–7 manifest the use of expressive language and lexical tropes in *поселян* (archaic word), *мобилизация* (metaphor), *нация* (metaphor), *военный стан* (metaphor). Syntactic foregrounding is observed in the use of mononuclear sentences and coordinate terms. In terms of Foregrounding in image, AOI 14 manifests object detalization and intense colour regime, while AOIs 15–17 do not.

Overall, 15 verbal mode cues of Graphic and orthographic foregrounding and 13 verbal mode cues of Linguistic foregrounding in text AOIs were attested. In image AOIs 11 pictorial mode cues of Foregrounding in image technique and 6 pictorial mode cues of Foregrounding in image colour were attested. To explore the effects of foregrounding cues onto the gaze behavior, the gaze measures Fixation Duration and Pupil Dilation were considered since the first is affected by information processing and

Fig. 1. AOI segmentation. Stimulus 6 (fragment)

the second with attention distribution [Chernigovskaya, Petrova (eds.), 2018; Prokofyeva, 2018; Kutlubaev et al., 2023]. Therefore, we identify the alignment patterns scaling the effects of semiotic foregrounding cues in text and pictorial stimuli areas of interest onto their gaze perception.

Experiment 2. Productive discourse task

In experiment 2 we directly determined the patterns of communicative alignment in speech and gesture. We collected film footages of multimodal behaviour elicited from 22 subjects (participants' written consents to use their image, video and behavioral data in scientific publications were obtained), transcribed and annotated for speech and gesture. The visual stimulus (VR-augmented dynamic scene based on themes of Van Gogh's "Starry Night" and "Bedroom in Arles") allows for information construal in three deictic dimensions: temporal deixis concomitant with narrative discourse passages, spatial deixis concomitant with descriptive discourse passages, and personal deixis concomitant with expository and argumentative discourse passages. During the productive phase of the experiment, the subjects (no longer exposed to the stimulus) were instructed to relate their watching experience to an interested partner. To identify the presence of a narrative, we used the discourse schemata of this type [Mandler, Johnson, 1977], which are setting, initiating events, characters' goals, attempts towards goals, and outcomes. Description is characterized by such discourse schemata as [MacSaveny, 2010] description/explanation, background information, elaboration, exemplification. Schemata typical of exposition warrant reason and come down to [Nippold, Scott (eds.), 2010] viewpoint formulation, viewpoint presentation, compare – contrast, cause – effect, and problem – solution. Argumentation necessitates justification with the following set of discourse schemata [van Eemeren, 2010]: standpoints at issue, starting points of discourse, argument advance, and outcome presentation. For instance, the narrative discourse schema ATTEMPTS TOWARDS GOALS can be manifested in verbs denoting movement, as in *потом мы* ***повернулись****, но вовнутрь мы* ***не возвращались*** // *afterwards* ***we turned around****, but* ***never got back*** *inside*; the use of a rhetorical question can introduce the expository schema VIEWPOINT FORMULATION as in ***Как они называются?..*** *ээ.. ну, спинки кровати...* // ***What do you call it?..*** *Hmm.. you know, the back of the bed...*

In terms of gesture, we considered four gesture functions as outlined in [Iriskhanova et al., 2023]: pragmatic (discourse structuring), representational (denoting shape, size, form), deictic (denoting direction), and adaptive (self-oriented movement). Figure 2 shows the co-occurrence of speech and gesture during the delivery of the discourse task by the subject.

We directly explore speech and gesture behavior (communicative multimodality) modulated by the discourse task of information transfer in monologue communication. Since narration is used to relate events, places, and characters, typically it is representational gestures that are found to be more frequent in accompanying narrative passages. We also expect adaptors to accompany instances of discourse hesitation frequently concomitant with relating expository and argumentative passages [Iriskhanova et al., 2023]. Therefore, in the present experiment we expect to scale discourse task effects onto gesture production specifying gesture types distribution.

Gesture:
deictic

Speech:
мы вылетаем из комнаты и движемся к небу / we're flying out of the room and are moving towards the sky

Fig. 2. Deictic gesture with the ATTEMPTS TOWARDS GOALS narrative discourse schema

Experiment 3. Receptive-productive discourse task

In experiment 3, we determined the patterns of perceptual alignment in visual and aural modes to a certain degree via the semiotic mode (areas of interest in the stimulus) and via communicative mode (speech produced on perceiving aurally presented information). The participants were engaged into a professional activity of remote simultaneous interpreting implying both comprehension and production [Keller et al., 2023] which is a highly demanding receptive-productive discourse task [Cienki, Iriskhanova, 2020; Gavrilenko, 2023]. Fourteen participants were subjected to the task with a visual stimulus present which adds an extra layer for information processing in interpreting [Yuan, Wang, 2023] (SMI red-x eye tracker was applied). The visual stimulus was a simulated popular science conference given via Zoom videoconference application with the overarching topic of "Green energy". Three areas of interest (AOIs) [Divjak, Milin, Medimorec, 2020] were identified on the screen: the speaker's head (for facial expressions), the Powerpoint presentation (for visual information, such as numbers, proper names, etc.) and the interpreters head, simulated via a small standing mirror placed in front of the monitor (Fig. 3).

Apart from the gaze duration in areas of interest in the visual stimulus, we attest post-hoc discourse modifications in speech (by contrasting the discourse structure of original text and interpreting text) following aurally/audially perceived information. To scale the aural effects in post-hoc discourse modifications in speech, we presume that aural perception is modulated by the productive discourse task the interpreter performs, which is preserving the information in discourse. Therefore, aural perception is explored indirectly via a contingent communicative mode of speech. Higher frequency of discourse modifications in speech signifies that aurally perceived information appears less significant for the interpreter. Applying W.L. Chafe's theory of "information packaging" [Chafe, 1976], we identified seven types of interpreter discourse modifications in speech, which can then be subdivided into two groups: the first comprises "Shift Changes": 1) "New versus Old"; 2) "Contrast Focus"; 3) "(Un)certainty"; 4) "Subject"; 5) "Topic as Context"; 6) "Perspective"; the second group includes: 7) "Omissions". In *…a variety of industrial applications,* ***like water desalination, enhanced oil recovery, food processing****, and so on and so forth = …может помогать* ***очищать воду от соли*** *и* ***в других целях*** – the interpreter, instead of listing the homogenous elements of an original phrase, relates one of the positions and then generalizes the rest. In ***This heat*** *– also known as* ***thermal energy****... = Эта* ***термальная энергия****, или* ***отопление****...* – the interpreter switches the order new and old information is presented in, which is supposedly done in order to evade a pause: this reversal served as a way for the participant to "find" the required equivalent in Russian.

Analyzing the eye and speech behaviour patterns leads us to hypothesizing a potential correlation between the three AOIs when it comes to gaze dwell time and speech modifications. Since we analyze contingency indirectly through two

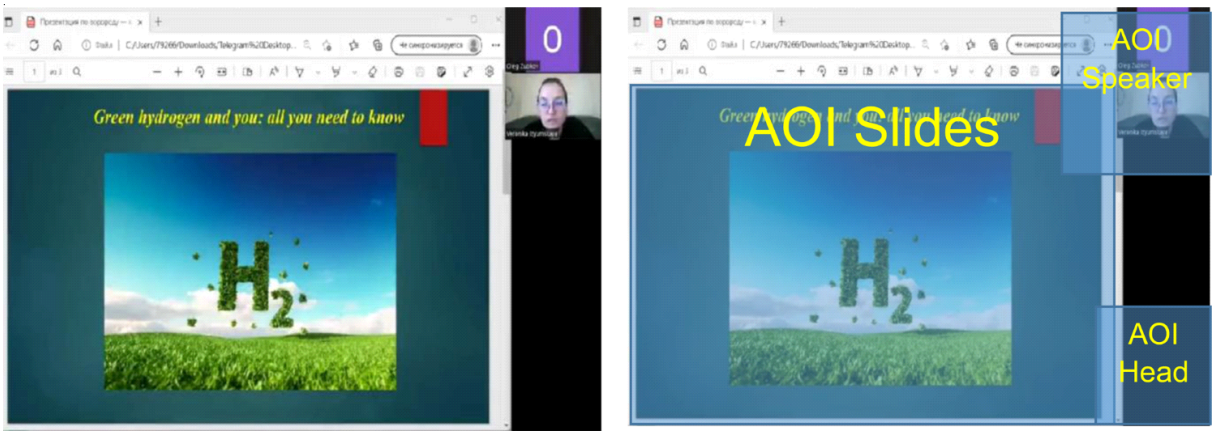

Fig. 3. AOI Segmentation

contingent modalities at once – through semiotic (AOIs) and communicative (speech production), a lower correlation result is to be expected.

Results and discussion

In this section, we present the results of three experimental studies determining the scales of crossmodal alignment of semiotic, communicative and perception modes in receptive, productive and receptive-productive discourse tasks.

Experiment 1. Receptive discourse task

To identify the foregrounding effects of each semiotic modality onto gaze perception, we used two gaze measures, pupil dilation known to be sensitive to perception foregrounding effects and fixation duration known to be sensitive to information retrieval effects [Chernigovskaya, Petrova (eds.), 2018; Kutlubaev et al., 2023]. We observed visible differences in pupil dilation with higher dilation in image areas and higher fixation duration in text areas. In Text, Average Pupil dilation is 2.78 mm, Average Fixation Duration is 159 ms. In Image, Average Pupil dilation is 2.95 mm, Average Fixation Duration is 152 ms.

To establish the effects of (1) verbal mode cues of Graphic and orthographic foregrounding and Linguistic foregrounding, (2) pictorial mode cues of Foregrounding in image technique and Foregrounding in image colour, onto pupil dilation and fixation duration, we performed one-way ANOVA tests separately for text and image AOIs (Table 2).

We observe a significantly larger number of graphic and orthographic foregrounding affecting pupil dilation which allows to presume that these cues relate to information perception rather that to information retrieval. A small number of linguistic foregrounding cues mediating pupil dilation suffices to presume that these cues do not produce steady increase on pupil diameter size since they do not mediate perception. Importantly, out of 11 Pictorial mode cues of Foregrounding in image technique and 6 cues of Foregrounding in image colour none was found to modulate fixation duration, whereas 4 cues of the first type affected pupil dilation. To scale the effects, we contrast the Kruskal – Wallis χ^2 and *p*-values and range them from higher (manifesting higher correlation) to lower. The scaled effects are established as follows:

Foregrounding in text via pupil dilation: Graphic and orthographic foregrounding

Foregrounding in image via pupil dilation: Foregrounding in image technique > Foregrounding in image colour

Foregrounding in text via fixation duration: Linguistic foregrounding

Foregrounding in image via fixation duration: None

Table 2. **Results of the one-way ANOVA (Foregrounding cues in text and image and gaze behavior)**

Gaze measures	Fixation Duration Kruskal – Wallis χ^2 [df; p]	Pupil Dilation Kruskal – Wallis χ^2 [df; p]
Foregrounding cues Graphic and orthographic foregrounding	use of brackets and inverted commas 5.17 [713; 0.023] first letter capitalization4.32 [713; 0.038]	use of italics 16.9 [713; <.001] use of brackets and inverted commas 4.27 [713; 0.039] capitalization of a word 3.65 [713; 0.05] non-standard orthography 4.55 [713; 0.033] use of interval 13.6 [713; <.001] exclamation or question mark 13.2 [713; <.001]
Linguistic foregrounding	use of proper names 6.33 [713; 0.012] use of expressive language and lexical tropes 3.63 [713; 0.057]	code-shifting 16.6 [713; <.001]
Foregrounding in image technique	None	collage technique 6.89 [625; 0.009] non-standard technics 6.17 [625; 0.013] non-standard stylistics 6.91 [625; 0.009] style modulation 4.52 [625; 0.033]
Foregrounding in image colour	None	classical triad colour spectrum 6.91 [625; 0.009] colour symbolism 8.97 [625; 0.003] colour effects 4.4 [625; 0.036]

The results manifest that while fixation duration is mostly susceptible to linguistic foregrounding, the pupil dilation responds to perception cues in both text and image. This serves an additional evidence of pupil dilation modulated by perception, attention distribution and general activation [Kutlubaev et al., 2023], and indirectly proves that gaze event duration responds to cognitive linguistic construal which underlies the use of linguistic foregrounding [Kiose et al., 2023].

Experiment 2. Productive discourse task

To identify the alignment patterns of speech and gesture manifested in discourse schemata in speech and functional gestures, we used direct identification regime. The results presented in Table 3 were processed with Pearson-correlation matrices and reveal the *r*-values (with *p*-values) of discursive effects in speech onto the use of gesture.

As expected, the analysis demonstrated the contingency of discourse schemata distribution on gesture use. These findings enabled us to establish the alignment patterns of speech and gesture by ranging Pearson's *r* and *p*-values from higher (manifesting higher correlation) to lower. The scaled effects are presented below:

Narration and gesture: representational > adaptive > deictic > pragmatic

Description and gesture: representational > adaptive > deictic > pragmatic

Exposition and gesture: representational > adaptive > pragmatic > deictic

Argumentation and gesture: adaptive > representational > pragmatic > deictic

We had predicted that there would be more representational gestures accompanying narrative and descriptive passages. Surprisingly, it also resulted in the most frequent type of gesture in expository passages; however, this could be explained by the fact that expository discourse markers were often embedded in narrative passages. Therefore, there was an ongoing overlap between representational gestures accompanying narration and embedded exposition. Figure 4 illustrates the coincidence of the narrative discourse marker expressed by the action verb *растягиваются / are stretched* and the embedded expository discourse marker expressed by the linker *тем не менее / nevertheless*.

Adaptors were the second most frequent gesture type in narrative, descriptive, and expository passages and the first most frequent one in argumentative passages. Nonetheless, it might be somewhat premature to make far-reaching conclusions about co-speech alignment here as its contingency upon speech functions was dismissed in [Iriskhanova et al., 2023].

Table 3. **Results of the correlation analysis (discourse schemata and gesture)**

Gesture types	Adaptive Pearson's *r* [df; p]	Pragmatic Pearson's *r* [df; p]	Representational Pearson's *r* [df; p]	Deictic Pearson's *r* [df; p]
Discourse schemata				
Narrative	0.713 [20; <.001]	0.611 [20; 0.003]	0.788 [20; <.001]	0.678 [20; <.001]
Descriptive	0.826 [20; <.001]	0.580 [20; 0.005]	0.889 [20; <.001]	0.736 [20; <.001]
Expository	0.765 [20; <.001]	0.738 [20; <.001]	0.857 [20; <.001]	0.730 [20; <.001]
Argumentative	0.763 [20; <.001]	0.750 [20; <.001]	0.752 [20; <.001]	0.554 [20; 0.007]

Gesture:
representational

Speech:
пропорции, **тем не менее**, сохранены, или в **какой-то** момент **растягиваются** [до дома, который находится перед нами] / the proportions are, **nonetheless**, consistent, or **at some point** are **stretched** [up to the house in front of us]

Fig. 4. The use of a representational gesture with narrative and expository discourse markers

Table 4. **Results of the correlation analysis (modifications in speech and AOI)**

Area of Interest	Slides Pearson's *r* [df; p]	Speaker Pearson's *r* [df; p]	Mirror Pearson's *r* [df; p]
Discourse Modification Type			
New versus Old	–0.177 [10; 0.582]	–0.056 [10; 0.864]	0.486 [10; 0.109]
Contrast Focus	–0.284 [10; 0.371]	0.556 [10; 0.060]	0.289 [10; 0.362]
(Un)certainty	–0.369 [10; 0.237]	0.101 [10; 0.755]	0.263 [10; 0.410]
Subject	–0.329 [10; 0.297]	0.248 [10; 0.437]	0.323 [10; 0.306]
Topic as Context	–0.562 [10; 0.057]	0.038 [10; 0.906]	0.595 [10; 0.041]
Perspective	–0.062 [10; 0.848]	0.547 [10; 0.066]	0.217 [10; 0.497]
Omissions	–0.232 [10; 0.468]	0.123 [10; 0.702]	–0.018 [10; 0.956]

Experiment 3. Receptive-productive discourse task

To identify the alignment patterns of perceptual modes, visual and aural, we used indirect identification regime. Visual mode was attested as contingent on a semiotic mode (fixation duration in 3 AOIs) while aural mode was explored via a contingent communicative mode of speech (discourse modifications). The results presented in Table 4 were processed with Pearson-correlation matrices, revealing the *r*-values (as well as *p*-values) of gaze effects in AOIs onto aural mode via a contingent communicative mode of speech.

The results demonstrate low contingency of discourse modification distribution on gaze behaviour. We presume that this is due to the fact that we analyzed contingency indirectly through two modalities at once – semiotic (AOIs) and communicative (speech production), thus it led to low values, which we had predicted. However, the results suggest the crossmodal explanation for the discrepancies in the use of speech and gesture in simultaneous interpreting found in [Cienki, Iriskhanova, 2020]. Nevertheless, scaling the significant or near-significant effects (ignoring the others, given their non-significant p-values) allows to obtain the following:

Visual perception via semiotic mode and Aural perception via discourse modifications in speech:
Slides AOI and modifications: Topic as Context
Speaker AOI and modifications: Contrast Focus
Mirror AOI and modifications: Topic as Context > New versus Old

The results manifest that "Topic as Context" modifications' p-values in AOIs of Powerpoint slides and the Mirror are close to 0.05 (0.057 and 0.041 respectively), Pearson's *r*-values are similar, but one is negative, while the other is positive. This observation allows to claim that modifications of topical information were seldom used in cases where participants paid close attention to the information presented on the screen, and conversely, were quite frequent during the periods of time when the participants looked at themselves in the mirror. The results suggest that enhanced visually and aurally perceived information leads to significantly lower modifications, whereas the lack of visual perception led to more modifications during the interpreting task.

Considering the results of three studies, we can identify the allowances and constraints of Crossmodal Alignment Framework. Communicative multimodality is subjected to direct observation [Iriskhanova, 2021]; consequently, the alignment patterns can manifest high contingency values (there is an overlap between gestures accompanying different speech patterns), which may mean that significant differences in the use of gestures with speech will be less frequently observed. Semiotic multimodality can be explored in receptive discourse tasks in its perception, which explains the efficiency of the studies determining the participants' gaze behavior in examining image and text semiotic multimodality [Chernigovskaya, Petrova (eds.), 2018; Kiose et al., 2023]; still the contingencies are less frequently observed and can be explained by the differences in perception. Semiotic alignment can also be identified via productive discourse tasks in communicating priorly visually perceived information, for instance in speech [Potekhin, 2023]; in this case contingencies will be expectedly lower being mediated by both perceptual and communicative modes. Perceptual multimodality can be explored as modulated by what people see (visually perceive) and what they relate in speech (audially perceive) – via the participants' gaze behavior

modulated by the areas of interest on the visual stimulus in the semiotic mode and via the participants' productive speech modulated by audially perceived information [Divjak, Milin, Medimorec, 2020]. In this case, the contingencies will be highly mediated by other modes input.

Conclusion

The study advances Crossmodal Alignment Framework which allows to explore the alignment patterns of semiotic, communicative and perceptual modes on common ground – via the discourse tasks, perceptive and productive. Assuming that the use of modes is controlled by cognitive growth points, the study takes this idea further to hypothesize that discourse tasks serve the cognitive growth points to determine the crossmodal alignment.

To attest to it, the study reports the results of three experiments which identify multimodal alignment patterns modulated by different discourse tasks – a receptive task in examining image and text, a productive task in discourse information construal using speech and gesture, and a receptive-productive task in visual and audial information perception and its further transfer in speech. In the study, we determined (1) the patterns of gaze behavior (fixation duration and pupil dilation) modulated by foregrounding in text and image, (2) the patterns of functional gesture modulated by discourse schemata of narration, description, exposition and argumentation, and (3) the patterns of gaze behavior in the stimuli areas of interest and discourse modification in speech of aurally perceived information. Additionally, we scaled the mode effects within each alignment pattern. Alignment of communicative modes (speech and gesture) appeared to manifest strong correlations, while alignment of semiotic modes explored via perceptual mode (gaze) produced lower contingencies and alignment of perceptual modes attested via semiotic and communicative modes displayed the lowest values. Although we found that indirect observation of mode alignment invariably leads to lower contingency effects of the modes, the results prove to be even more significant due to steady character of this contingency.

Overall, the study showed that examining the relations of different modes – semiotic, communicative, perception – is possible via direct and indirect observation. Crossmodal Alignment Framework allows both identifying their alignment patterns and scaling the contribution of each mode. Expectedly, the framework will allow to attest the input of each mode in operating different discourse tasks which is of particular importance in professional discourse and in mediated communication.

NOTE

[1] This research is part of the projects "Multimodal research of the speaker's communicative behavior in different discourse types" (075-03-2020-013) carried out at Moscow State Linguistic University and "Kinesic and vocal aspects of communication: parameters of variance" (FMNE-2022-0015) carried out at the Institute of Linguistics RAS.

[2] The samples taken from repositories served as the research data. The authors of the publication may not share the opinion of the authors of these open access internet samples.

REFERENCES

Chafe W.L., 1976. Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View. Li C.N., ed. *Subject and Topic*. New York, Academic Press, pp. 25-55.

Chernigovskaya T.V., Petrova T.E., eds., 2018. *The Gaze of Shroedinger's Cat: Identifying Gaze Metrics in Psycholinguistic Studies*. Saint Petersburg, Saint Petersburg University Publ. 228 p.

Cienki A., Iriskhanova O.K., 2020. Patterns of Multimodal Behavior Under Cognitive Load: An Analysis of Simultaneous Interpretation from L2 to L1. *Voprosy kognitivnoy lingvistiki* [Issues of Cognitive Linguistics], no. 1, pp. 5-11. DOI: https://doi.org/10.20916/1812-3228-2020-1-5-11

Divjak D., Milin P., Medimorec S., 2020. Construal in Language: A Visual-World Approach to the Effects of Linguistic Alternations on Event Perception and Conception. *Cognitive Linguistics*, no. 19, pp. 37-72. DOI: https://doi.org/10.1515/cog-2018-0103

Gavrilenko N.N., 2023. Sinteziruyushchiy etap v deyatelnosti otraslevogo perevodchika: determinatsiya otrasli i tekhnologiya formirovaniya neobkhodimykh kompetentsiy [Synthesizing Stage in the Activities of an

Industry Translator: Branch Determination and Technology for Forming Required Competences]. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Yazykoznanie* [Science Journal of Volgograd State University. Linguistics], vol. 22, no. 3, pp. 57-66. DOI: https://doi.org/10.15688/jvolsu2.2023.3.5

Iriskhanova O.K., 2014. *Igry fokusa v yazyke. Semantika, sintaksis i pragmatika defokusirovaniya* [Games of Focus in Language. Semantics, Syntax and Pragmatics Defocusing]. Moscow, Yaz. slav. kultury Publ. 320 p.

Iriskhanova O.K., 2021. *Polimodalnye izmereniya diskursa* [Multimodal Discourse Measures]. Moscow, Yaz. slav. kultury Publ. 448 p.

Iriskhanova O.K., Kiose M.I., Leonteva A.V., Agafonova O.V., Petrov A.A., 2023. Multimodal Collaboration in Expository Discourse: Verbal and Nonverbal Moves Alignment. *Lecture Notes in Artificial Intelligence*, vol. 14338, pp. 1-14. DOI: https://doi.org/10.1007/978-3-031-48309-7_29

Keller L., Viebahn M., Hervais-Adelman A., Seeber K., 2023. Unpacking the Multilingualism Continuum: An Investigation of Language Variety Co-Activation in Simultaneous Interpreters. *PLOS ONE*, vol. 18, no. 11, p. e0289484. DOI: https://doi.org/10.1371/journal.pone.0289484

Kiose M.I., Izmalkova A.I., Rzheshevskaya A.A., Makeev S.D., 2023. Sobytiye teksta i metateksta v glazodvigatelnom povedenii impulsivnykh i reflektivnykh chitateley [Text and Metatekst Event in the Gaze Behavior of Impulsive and Reflective Readers]. *Nauchnyy rezultat. Voprosy teoreticheskoy i prikladnoy lingvistiki* [Research Result. Theoretical and Applied Linguistics], vol. 9, no. 1, pp. 115-135. DOI: https://doi.org/10.18413/2313-8912-2023-9-1-0-8

Kutlubayev M.A., Shagieva D.R., Karimova G.I., Izmalkova A.I., Myachikov A.V., 2023. Pupillometriya v otsenke psikhoemotsionalnogo sostoyaniya i kognitivnykh funrtsiy cheloveks [Pupillometry in the Assessment of Emotional State and Cognitive Functions in Human]. *Zhurnal vysshey nervnoy deyatelnosti im. I.P. Pavlova* [I.P. Pavlov Journal of Higher Nervous Activity], vol. 73, no. 5, pp. 651-665. DOI: 10.31857/S0044467723050064

MacSaveny T., 2010. Towards a Description of Descriptive Discourse. *GIALens*, no. 3, pp. 1-5.

Mandler J.M., Johnson N.S., 1977. Remembrance of Things Parsed: Story Structure and Recall. *Cognitive Psychology*, vol. 9, no. 1, pp. 111-151.

McNeill D., 2006. Gesture, Gaze, and Ground. Renals S., Bengio S., eds. *Proceedings of Machine Learning for Multimodal Interaction: Second International Workshop 2005*. Berlin, Heidelberg, Springer Verlag, pp. 1-14.

Nippold M.A., Scott C.M., eds., 2010. *Expository Discourse in Children, Adolescents, and Adults: Development and Disorders.* New York, Psychology Press. 336 p.

Potekhin V.O., 2023. Diskursivnye skhemy i zhesty v immersivnoy kommunikatsii: vliyanie faktora tekhnologichnosti [Discourse Schemata and Gesture in Immersive Communication as Affected by Technology Factor]. *Cognitivnye issledovaniya yazyka* [Cognitive Studies of Language], vol. 5 (56), pp. 395-399.

Prokofyeva O.N., 2018. Spetsifika yazykovoy i zhestovoy peredatshi zritelnogo vospriyatiya [Visual, Verbal and Gestural Focusing in Russian Descriptive Discourse]. *Vestnik Moskovskogo gosudarstvennogo universiteta. Gumanitarnyye nauki* [Vestnik of Moscow State Linguistic University. Humanities], no. 18 (816), pp. 257-267.

Van Eemeren F.H., 2010. Strategic Maneuvering. Extending the Pragma-Dialectical Theory of Argumentation. *Argumentation in Context 2*. Amsterdam-Philadelphia, John Benjamins, pp. 93-127. DOI: https://doi.org/10.1075/aic.2

Van Leeuwen T., 2022. *Multimodality and Identity*. London, Routledge. 190 p.

Yuan L., Wang B., 2023. Cognitive Processing of the Extra Visual Layer of Live Captioning in Simultaneous Interpreting. Triangulation of Eye-Tracked Process and Performance Data. *Ampersand*, vol. 11, p. 100131. DOI: https://doi.org/10.1016/j.amper.2023.100131

Information About the Authors

Mariya I. Kiose, Doctor of Sciences (Philology), Chief Researcher, Centre for Socio-Cognitive Discourse Studies, Moscow State Linguistic University, Ostozhenka St, 38/1, 119034 Moscow, Russia; Leading Researcher, Laboratory of Multichannel Communication, Institute of Linguistics of the Russian Academy of Sciences, Bolshoi Kislovsky Lane, 1/1, 125009 Moscow, Russia, maria_kiose@mail.ru, https://orcid.org/0000-0001-7215-0604

Vadim O. Potekhin, Senior Lecturer, Diplomatic Academy of the Russian Ministry of Foreign Affairs, Ostozhenka St, 53/2, Bld. 1, 119021 Moscow, Russia; Postgraduate Student, Department of General and Comparative Linguistics, Moscow State Linguistic University, Ostozhenka St, 38/1, 119034 Moscow, Russia, v.potekhin@dipacademy.ru, https://orcid.org/0009-0003-2370-5097

Oleg D. Zubkov, Postgraduate Student, Department of General and Comparative Linguistics, Moscow State Linguistic University, Ostozhenka St, 38/1, 119034 Moscow, Russia, stabh7@gmail.com, https://orcid.org/0009-0008-9670-6273

Информация об авторах

Мария Ивановна Киосе, доктор филологических наук, главный научный сотрудник, Центр социокогнитивных исследований дискурса, Московский государственный лингвистический университет, ул. Остоженка, 38/1, 119034 г. Москва, Россия; ведущий научный сотрудник, Лаборатория мультиканальной коммуникации, Институт языкознания РАН, пер. Большой Кисловский, 1/1, 125009 г. Москва, Россия, maria_kiose@mail.ru, https://orcid.org/0000-0001-7215-0604

Вадим Олегович Потехин, старший преподаватель, Дипломатическая академия МИД России, ул. Остоженка, 53/2, стр. 1, 119021 г. Москва, Россия; аспирант кафедры общего и сравнительного языкознания, Московский государственный лингвистический университет, ул. Остоженка, 38/1, 119034 г. Москва, Россия, v.potekhin@dipacademy.ru, https://orcid.org/0009-0003-2370-5097

Олег Дмитриевич Зубков, аспирант кафедры общего и сравнительного языкознания, Московский государственный лингвистический университет, ул. Остоженка, 38/1, 119034 г. Москва, Россия, stabh7@gmail.com, https://orcid.org/0009-0008-9670-6273