Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference "Dialogue 2022"

Moscow, June 15–18, 2022

Gaze behavior in single-page monomodal and cross-modal switches as affected by Event construal

Maria Kiose
Moscow State Linguistic University,
Institute of Linguistics RAS, Moscow, Russia
maria_kiose@mail.ru

Anastasia Rzheshevskaya
Moscow State Linguistic University,
Moscow, Russia
arlen_nastya@rambler.ru

Anna Izmalkova
Moscow State Linguistic University,
Moscow, Russia
mayoran@mail.ru

Abstract

This paper proposes Event-modelling Framework (EMF) to explore the systemic Event construal effects onto gaze. In two eye tracking experiments examining monomodal and cross-modal switches with a total number of 898 gaze probes of Text and Picture Areas of Interest (AOIs), we investigated gaze behavior variations produced by Event construal registered in 61 parameters of Referent, Event Frame and Perspective construal. 5 non-aliased gaze metrics served to assess the significance of Event construal effects and scale them. The results contribute to the research field of ergonomic optimization in visual and reading tasks.

Keywords: gaze behavior, monomodal switches, cross-modal switches, Event-modelling Framework
DOI: 10.28995/2075-7182-2022-21-1078-1088

Глазодвигательное поведение при мономодальных и кросс-модальных переключениях в одностраничных стимулах. Влияние фактора конструирования событий

Мария Киосе
Московский государственный лингвистический университет,
Институт языкознания РАН,
Москва, Россия
maria_kiose@mail.ru

Анастасия Ржешевская
Московский государственный лингвистический университет,
Москва, Россия
arlen_nastya@rambler.ru

Анна Измалкова
Московский государственный лингвистический университет, Москва, Россия
mayoran@mail.ru

Аннотация

В работе предложена Модель организации события, позволяющая изучать системное влияние операций конструирования на изменения глазодвигательного поведения. В ходе двух окулографических экспериментов, исследующих мономодальные и кросс-модальные переключения, на материале 898 глазодвигательных проб в отношении зон интереса типа Текст и Изображение устанавливаются показатели сопряженности глазодвигательного поведения и 61 операции конструирования референта, событийной рамки и перспективы. Операции конструирования ранжируются по их значимости с учетом 5 глазодвигательных метрик. Исследование выполнено в сфере эргономики глазодвигательных усилий при выполнении задач, связанных с восприятием изображений и текстов.

Ключевые слова: глазодвигательное поведение, мономодальные переключения, кросс-модальные переключения, Модель организации события

1 Introduction

In recent years, eye-tracking research in cross-modal (for instance, in text and image) and monomodal switches has become an integral component of usability multimedia and VR studies. Gaze costs are explored as contingent on 1) foregrounding effects in pictorial modality [8, 33] or language modality [2, 8, 14, 30], 2) spatial effects in bimodal systems [18, 23, 25], 3) types of tasks performed [4, 12, 24], 4) several cognitive foregrounding effects, for instance dynamicity and agentivity [28, 35], 5) cross-modal information mapping [1, 30, 37]. Although prognostic computing has made significant progress, there are still challenges for assessing gaze effects in monomodal and bimodal systems as affected by systemic event construal. The aim of the work is to develop Event-modelling Framework (EMF) integrating operations of Referent, Event Frame and Perspective construal to investigate the gaze behavior as affected by differences in event construal. EMF is tested in two eye tracking experiments designed to assess event construal impact in single-page cross-modal and monomodal switches. The contributions of this study include (i) introducing EMF as potentially efficient in visual ergonomics; (ii) testing gaze behavior as affected by EMF operations in two experiments; (iii) scaling the EMF operations as producing higher gaze costs in observing Text and Image AOIs.

2 Event-modelling Framework

2.1 Related work

The proposed Event-modelling Framework is grounded in cognitive linguistics construal theories [22, 32] which have been implemented in experimental studies [7, 19], although the construal operations are scarcely if ever observed on a systemic basis. For instance, EMF was tested in eye tracking research in [10] where it allowed to determine the effects of space construal and agentivity in textual descriptions presented together with the images. In [21], the eye tracking procedure was applied to explore the effects of subjectivity in event construal as a component of EMF. It is noticeable that referent construal and event frame construal (event characteristics such as its dynamicity) effects are most tested in eye tracking psychological studies which are not attributed to EMF. For instance, it was experimentally verified that dynamic event construal patterns as well as referent visual and graphological salience produce higher gaze costs [2, 12, 26, 27]. Still, both psychological and cognitive linguistics works allow to consider three potential components of EMF which are Referent construal (in terms of its agentivity, singularity, among other characteristics), Event Frame construal (for example displayed in its completeness, manageability), and Perspective construal (which appears in subjectivity, observation point). Although they produce cumulative effects on the gaze costs, we still hypothesize (following the prior works given above) that we may observe their singular effects producing higher or lower gaze costs. To compile the lists of potential construal parameters, we addressed the theoretical works on EMF in cognitive linguistics. These parameters were further used to annotate the samples subjected to the eye tracking experiment.

Referent annotations were specified in J. Pustejovsky [29] who distinguishes four types of Referent construal operations or their "ways of seeing (WoS)", Part-whole, Kind, Functional, Life-history, as well as in A. Siewerska [31] who scales Referent accessibility in construal in terms of Speaker > addressee > non-participant, high physical salience > low physical salience, human > animate > non-animate (see also [32]). In oculographic studies Referent construal was most explored as regards its agentivity [2, 12, 27]; however, presumably other construal operations might affect gaze costs, for instance its physical salience or complexity. There have appeared several studies specifying the configurational characteristics [6, 25, 26], however they do not present a systemic view.

Event Frame annotations mostly describe event types which include achievements, accomplishments, states, and activities, and event characteristics like their dynamicity, aspectuality, repeatability, evaluation, entrenchment, and embodiment [9, 16, 29, 32]. In oculographic research the dynamic aspects of Event Frame construal appear most studied [3, 5], however, more recent studies explore event location and relations [10, 12, 18, 25].

Perspective annotations were proposed by A. Verhagen [36] who considers Perspective Coordination, Multiple Perspective Construal, Perspective Subjectivation, Perspective Shifts, and by O. Iriskhanova [15] who identifies the operations of Subjectivation and Objectivation, Vantage Point, Camera Angle, Scale, Coordinate System Configuration, Mental Scanning Path. Oculographic research of Perspective shifts is still uncommon; the work of Kleijn et al. [21] may serve a rare example of exploring the types of subjectivity in event construal.

2.2 Event-modelling annotations

In Appendix 1 we present EMF annotations following cognitive construal theories described above. Each Text and Image AOI is annotated and coded separately (101-119 for Referent construal, 201-228 for Event Frame construal, 301-314 for Perspective construal). Annotations are performed in UAM Image Tool (https://www.corpustool.com), which allows tiered annotations in single AOIs (see Figure 1). To process the data, we apply HETEROSTAT software [20] which allows to select single codes for further processing, tests the codes formatting validity, calculates parameter contingency (see Figure 2).

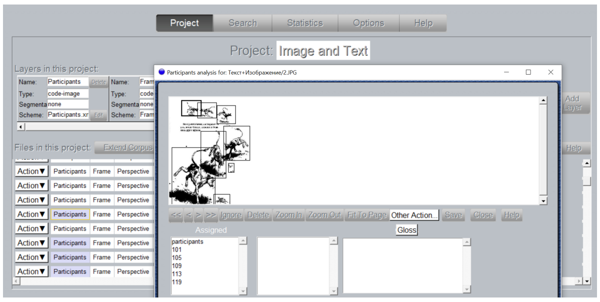


Figure 1: Annotating bimodal systems in UAM Image Tool

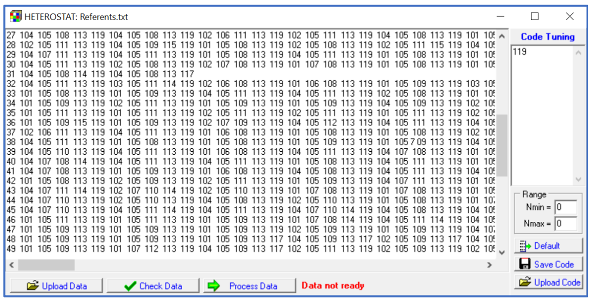


Figure 2: Processing the data in HETEROSTAT

Principal Component Analysis carried out in JAMOVI software (https://www.jamovi.org) allowed to select the samples manifesting different uniqueness indices (UI) for further eye tracking analysis.

3 Data and Methods

3.1 Design

Pre-experiment stimuli corpus. The manually annotated corpus consisted of 100 single-page drama play fragments and 100 single-page book fragments, which produced the UI range 0.145-0.727 for authors' text and 0.127-0.525 for participants' text in plays, and 0.096-0.676 for text and image event construal in book fragments. Due to higher range of UI in authors' text in drama plays, these texts with high and low UI were selected to be further submitted to eye tracking experiment. 4 text stimuli with UI equal to 0.145, 0.251, 0.327, 0.727 with total number of AOIs equal to 24 were used in Experiment 1. 6 book pages with UI equal to 0.369, 0.224, 0.2, 0.327, 0.51, 0.33 were selected for Experiment 2. They contained 46 AOIs, almost two times more AOIs than in the selected drama plays stimuli since we had to assess separately the gaze costs in Text AOIs and Image AOIs; therefore, the number of AOIs in Text-to-Text switches and Text-to-Image switches was compatible.

Eye-tracking experiments. In Experiment 1 testing monomodal switches, 4 text stimuli were presented in the same order for all subjects for a fixed time (360 seconds). 16 students (10 female, mean age = 23.2, age range = 19-28) participated in the experiment. 484 probes were subjected to further statistical analysis.

In Experiment 2 testing cross-modal switches, 15 students (10 female, mean age = 22, age range = 19-24) participated in the experiment. The experiment was organized as a single procedure, with each stimulus appearing for 15 seconds followed by 30 seconds inter-stimuli periods. We received a total number of Image AOI probes equal to 216, and Text AOI probes equal to 198.

In both experiments, the stimuli were presented on a 21-inch monitor. The SMI Red-x eye tracker was applied in the experiment (binocular system, frequency = 60 Hz, accuracy = 0.4°, head movement 40x20 cm, operating distance = 60-80 cm).

3.2 Stimuli

In Experiment 1 visual stimuli were 4 originally formatted pdf-texts, they were single-page fragments from drama plays “Music Classes” by L.S. Petrushevskaya (636 signs), “Biography” by L.N. Razumovskaya (1204 signs), “Reminiscences” by A.N. Arbuzov (730 signs), and “Dweller” by A.V. Vampilov (1373 signs). All 24 AOIs were visually salient, either given in bold type, in italics or in brackets. The average sign number (with spaces counted) in AOIs was 13.9 (range 4-47, SD = 10.7). To give an example of Event construal annotations, we will consider one AOI example.

Женский голос. Открой, детка, это я. ***Нина открывает дверь****. ‘Lady’s voice. Open, baby, open. It’s me.”* ***Nina opens the door****.’* (L.S. Petrushevskaya)

We observe the activity of Agentive participant (*Nina*), and Object/recipient (*door*), Single participant (*Nina*), Person-participant (*Nina*), Shifting/changing (*opens the door*), Spontaneous or occasional event, Single event, Interactive relations, True event, Manageable event, Author/speaker (the AOI is presented in the author’s text), Outer observer path (event presented by the author), Unidirectional and distinct path (there are no constraints to event observation).

In Experiment 2 there were 6 pdf single-page visual stimuli selected from Open-Access resources. The Image AOI Size range from Min=137475px to Max=8220575px, Mean=1360000px, Median=792872px. The Text AOI Size range from Min=245331px to Max=1140048px, Mean=150888px, Median=150888px. The number of signs in Texts AOIs varied from 5 to 186, SD = 61.1. To give an example of Event construal annotations, we will consider the AOI examples from Stimulus 6. It contains 10 AOIs which are shown in Figure 3.

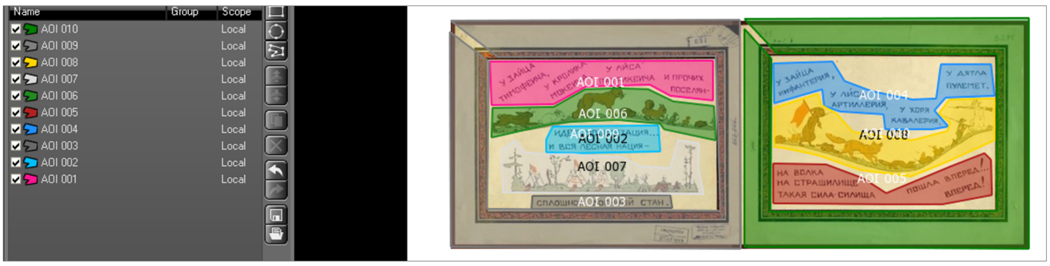


Figure 3: Sample 6 AOIs in SMI AOI Editor regime

Stimulus 6 displays 9 Agentive participants in Image AOIs, they are Animated. We observe the event from aside and all the Referents are Integral. The Referents are present both in Image and Text, they are also given in Text as Animated, Integral in *у зайца*, *у кролика, у лиса идет мобилизация*, the groups of Three and more participants / objects are present in *И вся лесная нация – сплошной военный стан (Ye. Venskii)*. The Event is of a Fictional type, this is a Shifting / changing event, it displays Interactive relations. Shifting / changing construal is duplicated in Image and in Text. It is also a Developing event which is manifested in parallel structures and process verbs *идет, пошла*. In Perspective construal we follow Outer observer path (we watch the event from aside) in both Image and Text.

All the Stimuli were annotated following the procedure described above. Each AOI received its annotations in terms of Referent, Event Frame and Perspective construal in binary format (1 for the presence of construal parameters, 0 for their absence). It allowed to perform multiple regression tests with the gaze metrics and parameter 1 and 0 values.

3.3 Data preparation

9 gaze metrics in AOIs [11, 17] were received; however, due to the presence of their aliased values we considered only 5 metrics, Dwell Time (DT), First Fixation Duration (FFD), Fixation Count (FC), Fixation Time (FT), and Average Fixation Duration (AFD). 15 regression tests in JAMOVI (since there were 3 Event construal groups and 5 gaze metrics) and 30 tests (15 with Text AOIs and 15 with Image AOIs) were performed with monomodal and cross-modal switches, respectively.

4 Results

4.1 Gaze metrics

In Figure 4 we present 5 gaze metrics in monomodal switches. Figure 5 shows the gaze metrics in Text AOIs in cross-modal switches.

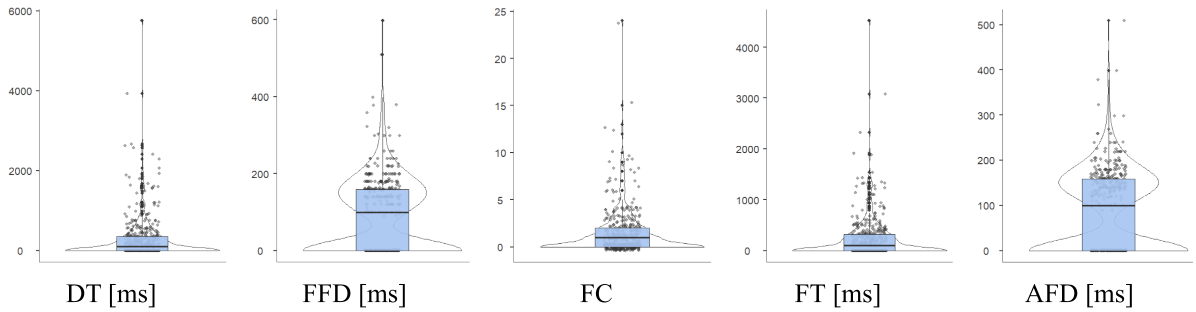


Figure 4: Gaze characteristics in monomodal switches

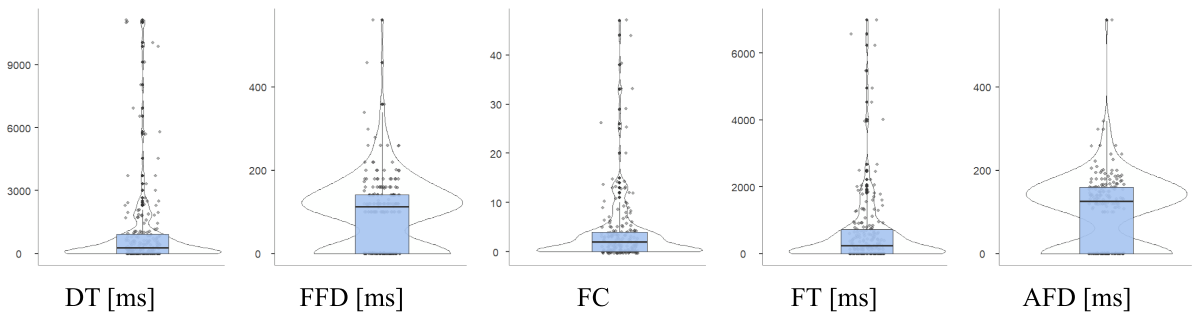
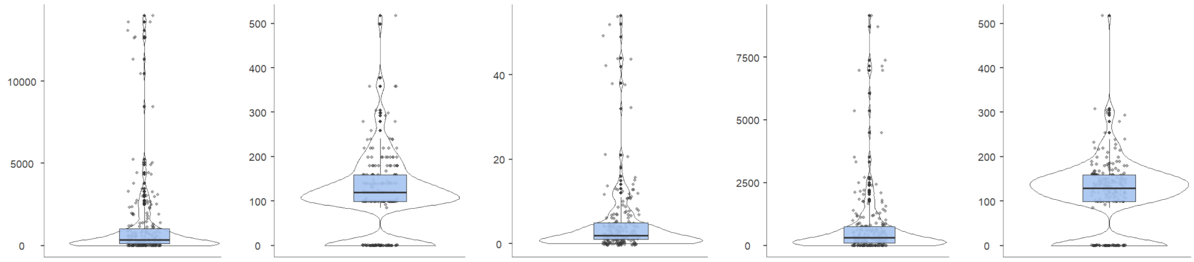


Figure 5: Gaze characteristics in cross-modal switches in Text AOIs

Both Figure 4 and Figure 5 manifest gaze metrics in Text AOIs with the difference in AOI sign number variance; however, we observe the contrasts which cannot follow this constraint. With N=391 in Experiment 1 and N=198 in Experiment 2, DT SD=565 and 2079, FC SD=2.5 and 7.43, FT SD=432 and 1147 may result from AOI sign number variance. At the same time, the differences which appear in FFD and AFD cannot be explained by higher AOI sign or size variance in cross-modal switches. In monomodal switches FFD Mean=88.4 and Median=99.4, whereas in cross-modal switches FFD Mean=100 and Median=112; additionally, in monomodal switches AFD Mean=85.6 and Median=89.5, whereas in cross-modal switches AFD Mean=105 and Median=125, which evidences in favor of higher gaze costs in bimodal systems Text AOI.

Figure 6 shows the gaze metrics in Image AOIs in cross-modal switches.



DT [ms] FFD [ms] FC FT [ms] AFD [ms]

Figure 6: Gaze characteristics in cross-modal switches in Image AOIs

With N=198 in Text AOIs and N=216 in Image AOIs, DT SD=2079 and 2550, FC SD=7.43 and 9.13, FT SD=1147 and 1464 cannot account for much higher AOI size variance (Text AOI Mean=150888px and Image AOI Mean=1360000px, almost 10 times higher) and presumably evidences of Image AOI lower gaze costs. It conforms to the results received in other eye tracking experiments testing variance in Text and Image AOIs [1, 7, 34], however it extends the data and specifies it in monomodal and cross-modal switches.

4.2 Gaze behavior as affected by Event construal in Text AOIs

In Table 1 we present the Regression models R^2 with 3 Event construal operations and 5 gaze metrics in Text AOI in Experiment 1. We specify the model predictors with significant p-values (*Bonferroni-Holm corrections indicated in case p exceeds $p_{\text{Bonferroni-Holm}}$).

Event construal	DT	FFD	FC	FT	AFD
Referent construal	0.192	0.108	0.195	0.187	0.093
	DT: Object / Recipient [4.47, < .001], Patient / Inactive participant [2.5, 0.013] **FFD:** Instrument [-2.57, 0.011], Single participant / object [2.36, 0.019] **FC:** Object / Recipient [4.1, < .001], Instrument [-2.14, 0.033], Patient / Inactive participant [2.78, 0.006] **FT:** Object / Recipient [4.15, < .001], Instrument [-2.04, 0.042], Patient / Inactive participant [2.51, 0.013] **AFD:** Single participant / object [2.37, 0.018]				
Event Frame construal	0.254	0.065	0.262	0.247	0.058
	DT: Shifting / changing [2.41, 0.017], Spontaneous or occasional event [1.93, 0.05], Event located in space [8.28, < .001] **FFD:** Perception [-1.96, 0.05, *0.025], Event located in space [2.09, 0.037, *0.025] **FC:** Shifting / changing [2.66, 0.008], Spontaneous or occasional event [2.37, 0.018], Event located in space [8.58, < .001] **FT:** Shifting / changing [2.73, 0.007], Spontaneous or occasional event [2.41, 0.016], Event located in space [8.07, < .001] **AFD:** Event located in space [2.939, 0.003]				
Perspective construal	0.153	0.018	0.149	0.142	0.022
	DT: Central area / elements [-5.7, < .001], Unidirectional and distinct path [4.75, < .001], Multidirectional but distinct path [2.66, 0.008], Regular key participant [-2.13, 0.034], Central event [2.18, 0.03] **FFD:** None **FC:** Central area / elements [-5.47, < .001], Unidirectional and distinct path [4.21, < .001], Multidirectional but distinct path [2.05, 0.041] **FT:** Central area / elements [-5.43, < .001], Unidirectional and distinct path [4.19, < .001], Multidirectional but distinct path [2.09, 0.038], Regular key participant [-1.97, 0.049] **AFD:** None				

Table 1: Monomodal switches in Text AOIs.
Regression models R^2 and best predictors: t, p, *$p_{\text{Bonferroni-Holm}}$

In Table 2 we present the Regression models R^2 with 3 Event construal operations and 5 gaze metrics in Text AOI in Experiment 2.

<table>
<tr><th>Event construal</th><th>DT</th><th>FFD</th><th>FC</th><th>FT</th><th>AFD</th></tr>
<tr><td rowspan="2">Referent construal</td><td>0.74</td><td>0.226</td><td>0.723</td><td>0.704</td><td>0.249</td></tr>
<tr><td colspan="5">DT: Abstract object [16.84, < .001]
FFD: Object / Recipient [-2.49, 0.014], Patient / Inactive participant [-2.47, 0.014], Two participants / objects [2.24, 0.027], Three and more participants / objects [-2.01, 0.046, *0.035], Animated participant [2.0, 0.047, *0.035], Inanimate participant / object [2.06, 0.041, *0.035], Integral participant / object [2.06, 0.041, *0.035], Facial movements of participant [-3.32, 0.001], Author / Speaker [-2.63, 0.009]
FC: Abstract object [17.19, < .001]
FT: Abstract object [16.53, < .001]
AFD: Object / Recipient [-2.4, 0.017], Patient / Inactive participant [-2.6, 0.01], Facial movements of participant [-3.0, 0.003], Author / Speaker [-3.0, 0.003]</td></tr>
<tr><td rowspan="2">Event Frame construal</td><td>0.405</td><td>0.183</td><td>0.376</td><td>0.351</td><td>0.197</td></tr>
<tr><td colspan="5">DT: True or real event [5.15, < .001], Shifting / changing [-8.82, < .001], Spontaneous or occasional event [-5.26, < .001], Flash event [-6.1, < .001]
FFD: None
FC: True or real event [3.73, < .001], Shifting / changing [-8.93, < .001], Spontaneous or occasional event [-4.03, < .001], Flash event [-6.45, < .001]
FT: True or real event [3.19, 0.002], Shifting / changing [-8.51, < .001], Spontaneous or occasional event [-3.43, < .001], Flash event [-6.01, < .001]
AFD: None</td></tr>
<tr><td rowspan="2">Perspective construal</td><td>0.405</td><td>0.183</td><td>0.376</td><td>0.351</td><td>0.197</td></tr>
<tr><td colspan="5">DT: Unidirectional and distinct path [-6.09, < .001]
FFD: None
FC: Outer observer path [2.15, 0.033], Unidirectional and distinct path [-6.55, < .001], Multidirectional but distinct path [-2.11, 0.037]
FT: Outer observer path [2.01, 0.046], Unidirectional and distinct path [-6.28, < .001], Multidirectional but distinct path [-2.37, 0.019]
AFD: None</td></tr>
</table>

Table 2: Cross-modal switches in Text AOIs. Regression models R^2 and best predictors: t, p, $*p_{\text{Bonferroni-Holm}}$

Model predictability in Text AOIs in cross-modal switches is higher as compared to monomodal switches. Since three gaze metrics, DT, FC, and FT display high variability due to AOI size and sign number variance, their regression model values are less reliable, although they manifest the general tendencies. At the same time, the regression model values describing FFD and AFD appear more significant. We may observe that this is mostly Referent construal that affects FFD and AFD in Text AOIs in both monomodal and bimodal systems.

In the experiment, in monomodal switches FFD and AFD are affected by Single participant / object and Event located in space; in cross-modal switches FFD is affected by Two participants / objects. Object / Recipient, Patient / Inactive participant, Facial movements of participant, Author / Speaker lower the gaze costs. These results do partially correlate with the results reported in [3, 4, 5, 23, 25]; however, we revealed that dynamicity, interactionality, emotionality produce steady negative and positive gaze effects in Text AOIs in cross-modal switches, whereas in monomodal switches referent prominence and space location were most counted. Additionally, fictionality revealed in [6, 27] is not found as potent in the attested stimuli. Perspective construal appears highly predictive in cross-modal switches in Text AOIs (here displayed in Author / Speaker), which correlates with the results reported in Kleijn et al. [21]. Importantly, R^2 with FFD and AFD in Text AOIs is higher in cross-modal switches, which means that several event construal operations in Text AOIs in bimodal systems produce steadier gaze effects.

4.3 Gaze behavior as affected by Event construal in Image AOIs

In Table 3 we present the Regression models R^2 with 3 Event construal operations and 5 gaze metrics in Image AOI in Experiment 2.

<table>
<tr><th>Event construal</th><th>DT</th><th>FFD</th><th>FC</th><th>FT</th><th>AFD</th></tr>
<tr><td rowspan="2">Referent construal</td><td>0.886</td><td>0.092</td><td>0.843</td><td>0.819</td><td>0.07</td></tr>
<tr><td colspan="5">DT: Patient / Inactive participant [4.47, < .001], Single participant / object [4.77, < .001], Two participants / objects [-6.4, < .001], Three and more participants / objects [2.26, 0.025], Person-participant: [-5.28, < .001], Animated participant [-4.99, < .001], Inanimate participant / object [-4.69, < .001], Abstract object [-5.31, < .001], Integral participant / object [4.75, < .001], Author / Speaker [5.65, < .001]
FFD: Patient / Inactive participant [2.34, 0.02, *0.016], Two participants / objects [-2.32, 0.021, *0.016], Author / Speaker [2.06, 0.041, *0.016]
FC: Patient / Inactive participant [3.12, 0.002], Single participant / object [3.7, < .001], Two participants / objects [-4.75, < .001], Person-participant [-4.04, < .001], Animated participant [-3.72, < .001], Inanimate participant / object [-3.34, < .001], Abstract object [-3.94, < .001], Integral participant / object [3.35, < .001], Author / Speaker [4.29, < .001]
FT: Patient / Inactive participant [2.74, 0.007], Single participant / object [3.1, 0.002], Two participants / objects [-4.19, < .001], Person-participant [-3.43, < .001], Animated participant [-3.16, < .001], Inanimate participant / object [-2.95, < .001], Abstract object [-3.38, < .001], Integral participant / object [2.92, 0.004], Author / Speaker [3.71, < .001]
AFD: None</td></tr>
<tr><td rowspan="2">Event Frame construal</td><td>0.846</td><td>0.046</td><td>0.815</td><td>0.796</td><td>0.018</td></tr>
<tr><td colspan="5">DT: Creation / destruction [-26.72, < .001], Shifting / changing [-31.28, < .001], Perception [-25.12, < .001], Interpersonal relations [7.05, < .001]
FFD: None
FC: Creation / destruction [-23.91, < .001], Shifting / changing [-27.8, < .001], Perception: [-22.23, < .001], Interpersonal relations [6.16, < .001]
FT: Creation / destruction [-22.4, < .001], Shifting / changing [-26.32, < .001], Perception [-20.61, < .001], Interpersonal relations [5.28, < .001]
AFD: None</td></tr>
<tr><td rowspan="2">Perspective construal</td><td>0.819</td><td>0.027</td><td>0.787</td><td>0.772</td><td>0.017</td></tr>
<tr><td colspan="5">DT: Outer observer path [30.14, < .001], Unidirectional and distinct path [3.26, 0.001]
FFD: None
FC: Outer observer path [27.21, < .001], Unidirectional and distinct path [2.32, 0.022]
FT: Outer observer path [25.97, < .001]
AFD: None</td></tr>
</table>

Table 3: Cross-modal switches in Image AOIs.
Regression models R^2 and best predictors: t, p, $*p_{\text{Bonferroni-Holm}}$

The results show that R^2 in Image AOIs is even higher than in Text AOIs, however the values describing FFD and AFD are on the opposite much lower, additionally, there are no good predictors of these gaze metrics. Overall, we revealed 1) several predictors of FFD and AFD (more reliable gaze metrics) and multiple predictors of less reliable metrics, 2) scaled model prognostic values and the values of single operations of EMF as gaze predictors in a) monomodal and bimodal systems, b) Text and Image AOIs. Event Frame construal operations predicting lower gaze costs in Image AOIs (revealed as tendencies) are Creation / destruction, Shifting / changing, Perception; their estimates are high, which was much unexpected, since these are the dynamic events that are generally reported to produce higher gaze effects [12, 28, 35]. However, we presume that this results from distributed Event construal in Text and Image AOIs; in all probability Perspective construal in Image AOIs with higher gaze costs alleviates gaze costs in Event Frame construal, which proves that further research in Perspective construal and specifically in subjectivity initiated in [21] is necessary. Perspective construal in Text AOIs although does not stimulate higher gaze costs in the attested stimuli. We observe their higher values in Abstract object construal, which was expected as abstract referents are less salient (or less accessible) as stated in theoretical studies on Referent construal [31], therefore the study supports this view. It also necessitates further research in the sphere of abstract and bodily referents construal which is also discussed in [2, 8], and mostly explored in metaphor studies [13, 14] and embodiment studies [26], however it is less considered in gaze cost studies as opposed to static and dynamic referents [12, 27].

5 Final remarks

In the paper, we proposed Event-modelling Framework to explore the gaze costs as potentially stimulated by event construal effects. Although the efficiency of EMF has been preliminary attested in several eye tracking studies [10, 21] and several event construal effects have also been explored in cognitive psychological frameworks [2, 12, 26, 27], there is still a need to observe 1) their cumulative effects and 2) the effects of single construal parameters as contrasted with the effects of other parameters. Therefore, we develop a three-component Event model Framework comprising Referent, Event frame and Perspective construal accommodating 61 parameters, described by event construal operations. We then tested it to explore the gaze effects in single-page monomodal and cross-modal switches in text and image areas of interest. To explore the event construal effects, we used regression modelling applying Event construal operations and 5 gaze metrics in Text and Image AOI in two experiments contrasting their results. The results show that Model predictability in Text AOIs in cross-modal switches is higher as compared to monomodal switches. Importantly, we found out that it was Perspective construal which was highly predictive in Image AOIs in cross-modal switches. In monomodal switches in Text AOIs, the gaze costs were higher in case the Referent was Object / Recipient and the Event was located in space and was of a Shifting / Changing and Spontaneous or occasional type. In cross-modal switches in Text AOIs, the gaze costs were higher in case the Referent was of Abstract type and the Event was True or real. In cross-modal switches in Image AOIs, very high gaze costs were produced by Outer observer path in Perspective construal (when there was no inner observer), when there were Interpersonal relations in Event Frame construal, when the Referent was Single, Integral, when the Author or Speaker was present in the Image. The results although received on a limited number of stimuli allow to predict possible clines in monomodal and cross-modal switches as affected by event construal in the stimuli. The data presenting the results of both experiments were made publicly available[1].

Overall, Event-modelling Framework allows to study multimodal systems construal since different modalities can be explored on the same grounds, still it is similarly potent in monomodal systems analysis. EMF appears potentially effective in ergonomic optimization of visual Tasks. When developed for particular types of stimuli, it may allow to predict what will attract the viewers' and readers' attention or what will escape their attention depending on the event type and its components. At the same time, the study has shown that there exist several constraints in EMF application, for instance collinearity effects, low regression model predictability, AOI size gaze metrics dependency.

Acknowledgements

This research is financially supported by the Russian Science Foundation, project No. 22-28-01754 "Cognitive load economy in media texts interpretation: Multimodal Corpus of Oculographic Reactions MultiCOR".

References

[1] Cengiz Acarturk, Christopher Habel, Kursat Cagiltay and Özge Alaçam. Multimodal comprehension of language and graphics: Graphs with and without annotations. *Journal of Eye Movement Research*, 1(3), 2007. P. 1–15.

[2] Gery T.M. Altmann and Yuki Kamide. Discourse-mediation of the mapping between language and the visual world: eye-movements and mental representation. *Cognition*, 111, 2009. P. 55–71.

[3] Vladimir A. Barabanshhikov. *Facial expressions and their perception*. Moscow: RAS Institute of Psychology, 2012.

[4] Elena Carniglia, Marcella Caputi, Valentina Manfredi, Daniela Zambarbieri and Eliano Pessa. The influence of emotional picture thematic content on exploratory eye movements. *Journal of Eye Movement Research,* 4(4), 2012. P. 1–9.

[5] Siyuan Chen and Julien Epps. Automatic classification of eye activity for cognitive load measurement with emotion interference. *Computer Methods and Programs in Biomedicine*, 110(2), 2013. P. 111–124.

[6] Viviane Clay, Peter König and Sabine U. König. Eye tracking in virtual reality. *Journal of Eye movement research,* 12, 2019. https://doi.org/10.16910/jemr.12.1.3

[1] The data are available in the following OSF repository: https://osf.io/r43pa/files/

[7] Neil Cohn and Martin Paczynski. The neurophysiology of event processing in language and visual events. In: Robert Truswell (ed.) *The Oxford handbook of event structure.* Oxford: Oxford University Press, 2019. P. 623–637.
[8] Kenny R. Coventry, Dermot Lynott, Angelo Cangelosi, Lynn Monrouxe, Dan Joyce and Daniel C. Richardson. Spatial language, visual attention, and perceptual simulation. *Brain and Language,* 112(3), 2010. P. 202–213.
[9] Valery Z. Demyankov. Event in semantics, pragmatics and interpretation coordinates. *Izvestiya AN USSR. Literature and language series,* 42(4), 1983. P. 320–329.
[10] Dagmar Divjak, Petar Milin and Srdan Medimorec. Construal in language: A visual-world approach to the effects of linguistic alternations on event perception and conception. *Cognitive Linguistics,* 19, 2020. P. 37–72.
[11] Andrew Duchowski. *Eye Tracking Methodology: Theory and Practice*. London, Springer, 2007.
[12] Monique Flecken, Johannes Gerwien, Mary Carroll and Christiane von Stutterheim. Analyzing gaze allocation during language planning. A cross-linguistic study on dynamic events. *Language and Cognition,* 7, 2015. P. 138–166.
[13] Raymond W. Gibbs Jr. and Marcus Perlman. 2006. The contested impact of cognitive linguistic research on the psycholinguistics of metaphor understanding In: Gitte Kristiansen, Michel Achard, Rene Dirven and Francisco J. Ruiz Mendoza Ibáñez (eds.) *Cognitive linguistics: Current applications and future perspectives*. Berlin: Mouton de Gruyter, 2006. P. 211–228.
[14] Shir Givoni and Rachel Giora. Salience and Defaultness. In: Frank Liedtke and Astrid Tuchen (eds.) *Handbuch Pragmatik.* Stuttgart: J.B. Metzler, 2018. P. 207–213.
[15] Olga K. Iriskhanova. On perspectivization in cognitive linguistics. *Cognitive Studies of Language,* 15, 2013. P. 43–58.
[16] Ray Jackendoff. *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press, 2002.
[17] Roger Johansson, Franziska Oren and Kenneth Holmqvist. Gaze patterns reveal how situation models and text representations contribute to episodic text memory. *Cognition, 175*, 2018. P. 53–68.
[18] Cheryl Johnson and Richard Mayer. 2012. An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology: Applied, 18(2)*, 2012. P. 178–191.
[19] Maria I. Kiose and Anastasia A. Rzheshevskaya. Interdiscourse switching in drama text. Gaze patterns and comprehension check results. *Issues in Cognitive Linguistics, 4,* 2021. P. 55–66.
[20] Maria I. Kiose and Andrey A. Efremov. HETEROSTAT software. Registered in FIPI 21.09.2020, registration number 2020661240.
[21] Suzanne Kleijn, Willem M. Mak and Ted J. M. Sanders. Causality, subjectivity and mental spaces: Insights from on-line discourse processing. *Cognitive Linguistics, 32(1)*, 2021. P. 35–65.
[22] Ronald W. Langacker. Construal. In: Ewa Dąbrowska and Dagmar Divjak (eds.) *Handbook of Cognitive Linguistics.* Berlin and Boston: De Gruyter Mouton, 2015. P. 120–143.
[23] Jesus Mayor, Laura Raya and Alberto Sanchez. A comparative study of virtual methods of interaction and locomotion based on presence, cybersickness and usability. *IEEE Trans. Emerg. Topics Comput. 9*, 2021. P. 1542–1553.
[24] Marino Menozzi, A von Buol, Helmut Krueger and C Miège. Direction of gaze and comfort: Discovering the relation for the ergonomic optimization of visual Tasks. *Ophthalmic and Physiological Optics,* 14(4), 1994. P. 393–399.
[25] Ya-feng Niu, Hong-rui Zuo, Xin Yang, Cheng-qi Xue, Ning-yue Peng, Lei Zhou, Xiao-zhou Zhou and Tao Jin. 2021. Improving accuracy of gaze-control tools: Design recommendations for optimum position, sizes, and spacing of interactive objects. *Human Factors and Ergonomics in Manufacturing and Service Industries,* 31(3), 2021. P. 249–269.
[26] Alva Noë and Kevin J. O'Regan. A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences,* 24(5), 2001. P. 883–917.
[27] Anna Papafragou, Justin Hulbert and John Trueswell. Does language guide event perception? Evidence from eye movements. *Cognition,* 108, 2008. P. 155–184.
[28] Frank Papenmeier and Markus Huff. DynAOI: A tool for matching eye-movement data with dynamic areas of interest in animation and movies. *Behavior research methods,* 42(1), 2010. P. 179–187.
[29] James Pustejovsky. *The Generative Lexicon*. Cambridge, Massachusetts: MIT Press, 1995.
[30] Keith Rayner. The 35th Sir Frederick Barlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quaterly Journal of Experimental Psychology,* 62, 2009. P. 1457–1506.
[31] Anna Siewerska. On the discourse basis of person agreement. In: Tuija Virtanen (ed.) *Approaches to cognition through text and discourse*. Berlin, New York: Mouton de Gruyter, 2004. P. 33–48.
[32] Leonard Talmy. *Toward a Cognitive Semantics*. Two volumes. Cambridge: MIT Press, 2000.

[33] Pavel Ugwitz, Ondřej Kvarda, Zuzana Juříková, Čeněk Šašinka and Sascha Tamm. Eye-tracking and interactive virtual environments: Implementation and Evaluation. *Applied Science,* 12(3): 1027. 2022.
[34] Geoffrey Underwood and Ralph Radach. Eye guidance and visual information processing: Reading, visual search, picture perception and driving. In: Geoffrey Underwood (ed.) *Eye guidance in reading and scene perception*. Oxford: Elsevier, 1998. P. 1–28
[35] Boris M. Velichkovsky, Marcus Joos, Jens R. Helmert and Sebastian Pannasch. Two visual systems and their eye movements: Evidence from static and dynamic scene perception. In: Bruno Bara, Larry Barsalou and Monica Bucciarelli *Proceedings of the XXVII Conference of the Cognitive Science Society.* Mahwah, NY: Lawrence Erlbaum, 2005. P. 2283–2288.
[36] Arie Verhagen. Construal and perspectivization. In: Dirk Geeraerts and Hubert Cuyckens (eds.) *The Oxford Handbook of Cognitive Linguistics*. Oxford: Oxford University Press, 2007. P. 48–81.
[37] Tatiana A. Zlokazova, Irina V. Blinnikova, Sergei S. Grogorovich and Ivan V. Burmistrov. Search Results on Flight Booking Websites: *Displaying Departure and Return Flights on a Single Page vs Two Consecutive Pages*. In: D. Lamas et al. (eds.) INTERACT 2019, LNCS 11749, 2019. P. 668–671.

Appendix 1. Event-modelling Framework annotations

Event Construal Type	**Operation Group**	**Operation**
Referent	Agentivity	Agentive participant / Object or Recipient / Instrument / Patient or Inactive participant
	Number	Single participant or object / Two participants or objects / Three and more participants or objects
	Reference	Person-participant / Animated participant (e.g., animal) / Animated object / Inanimate participant or object / Abstract object
	Referential integrity	Integral participant or object / Component of participant or object / Head of participant / Facial movements of participant
	Personalization	Author or Speaker / Addressee or Reader / Unidentified participant
Event Frame	Truth	True or real event / Fictious event
	Type	Creation or destruction / Shifting or changing / Perception / Lack of action or dynamics
	Relations	Interpersonal relations / Interactive relations / No interpersonal or interactive relations
	Manageability	Spontaneous or occasional event / Manageable event
	Completeness	Completed event / Unfinished event
	Instantness	Flash event / Developing event
	Achievement	Event-achievement / Absence of event achievement
	Evaluation	Positive evaluation / Negative evaluation / Event lacking evaluation
	Space location	Event located in space / Event with no space location
	Time location	Event located in time / Event with no time location
	Repeatability	Repeated event / Single event
	Cause and effect	Event with cause and effect stated / Event without cause and effect stated
Perspective	Vantage Point	Outer observer path / Inner observer path
	Viewpoint	Central area or elements / Peripheral area or elements
	Distancing	Zooming out / Zooming in / Distancing shifts
	Observation Path	Unidirectional and distinct path / Multidirectional but distinct path / Vague path
	Key Participant Centrality	Regular key participant / New key participant
	Event Centrality	Central event / Secondary event