NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS

As a manuscript

Mikhail Pokhoday

Modality of attention and syntactic choice in English and Russian languages

PhD Dissertation Summary
for the purpose of obtaining academic degree
Doctor of Philosophy in Psychology HSE

Academic supervisor:
Andriy Myachykov

PhD, Professor

Moscow, 2020

Table of contents

Introduction

It is a well-established fact that manipulation of attention using visual priming affects assignment of grammatical roles in the produced sentence in various languages (Tomlin, Pokhoday and Myachykov, 2018). Yet, manipulation of attention is not only a visual modality prerogative. Our brain simultaneously processes input from multiple sensory modalities including auditory, olfactory, tactile etc. Attention works as a filter picking up what is necessary in each moment for a given task. Thus, it is possible and there is some evidence that priming from other attentional modalities (auditory, motor, and/or multimodal) might have similar to visual if not stronger effects on syntactic choice. The aim of this thesis was to evaluate how and if priming of attention by visual, auditory, motor, and multimodal cues affects syntactic choice. In this project native English and Russian speakers described cartoon-like depictions of transitive events between animated and non-animated entities after their attention was primed by explicit, non-informative cue/cues towards one of the interacting protagonists/objects. We used eye tracking to control allocation of attention, yet the eye tracking data will be discussed in later papers. The aim of this work was to extend the already existing knowledge about the language production mechanisms, in particular how attention and syntax interact during sentence production. There were two main questions that drove current work: knowing that visual attention affects the structure of the produced sentences it is unknown whether auditory and motor directional explicit non-informative attentional cues (or a combination of visual, auditory and motor cues) produce similar effects to visual primes. Secondly, it was interesting to see if there are differences or similarities in syntactic choice during sentence production between languages with strict (English language) and free (Russian language) word order languages. In addition to the above questions, we wanted to investigate how orientation of the event left to right or right to left might influence the assignment of grammatical roles by speakers of Russian language. Let us start with an overview of theoretical and experimental literature on the discussed topic.

To begin with, we need to consider how attention and conceptualisation of the described event are explained in commonly accepted theoretical frameworks of sentence production. One of the most enduring models (Levelt, 1989) starts with a conceptualiser that allows mapping of a conceptual representation onto a pre-verbal *message*. It is during this stage that the speaker constructs the pre- or non-linguistic *message*, which serves as input to an *utterance formulator*. It is this message that carries forward the conceptual formulation of the event to be described including its perceptual map together with the salience of its components. Then the utterance formulator proceeds by selecting *lemmas*, or the words with their grammatical properties, and producing a *structural assembly* – prescribing structural configuration and a word order according to the rules of a given grammar. Finally, the articulator operates on this input producing spoken (or written) linguistic output. The stages that follow the message apprehension are called *linguistic stages* as operations within them have to do with lexical, morphological, syntactic, and phonological access. Levelt's model is especially good in detailing how these linguistic stages work while it is quite general about the organisation of the message (or, indeed, whether the message itself has any internal organisation) and the mechanisms of mapping the content of the message onto the corresponding linguistic categories. In order to minimally describe how the content and organization of the message may affect the content and the organization of a sentence, we need to consider two very important theoretical contrasts that we will return to when we discuss the data. To situate these contrasts in context, consider an event, in which a boy is kissing a girl. In English, this same event can be described by using multiple structural alternatives including the following:

(a) *The boy is kissing a girl.*
(b) *The girl is kissed by the boy.*
(c) *It is the girl that the boy is kissing.*

Although these do not fully exhaust available structural inventory, they allow us to identify three distinct structural scenarios:

(a) An active clause with *the boy* as syntactic <u>subject</u> and the sentential <u>starting point</u> with *the girl* as <u>object</u> in <u>final position</u>.

(b) A passive clause where *the girl* is a syntactic subject and the sentential starting point, and *the boy* is a part of a final position adverbial.
(c) A cleft sentence with *the girl* being a part of a main clause (a complement object) and the *boy* as a subject of the postposed embedded complement clause.

At this point it clearly seen that that a speaker must go through some decision-making process (explicit or implicit) and to choose between those structures. There are some theoretical proposals that proclaim that perception might influence this process. Representations of objects in the human mind and locations are mapped on the distinction between nouns and prepositions according to some theories (Landau and Jackendoff, 1993; Jackendoff, 1996). Also, existing literature provides (1) structural, (2) developmental, and (3) behavioural reasons to explain how speakers decide between described scenarios and how produced language relies upon allocation of attentional resources. Further, I will review the studies that demonstrate that attention and language are finely intertwined.

The structural experimental evidence suggest that various human brain regions are usually involved in a few mental operations. Neuroimaging studies in reading, for example, have successfully identified brain regions controlling the process of chunking of visual letters into words, connecting letters with sounds, and accessing a distributed lexicon of semantics. According for example to McCandliss and colleagues (2003) in the process of chunking a posterior visually specific area of the left fusiform gyrus is involved. Interestingly, similar areas in the right hemisphere are involved in the facial processing (Kanwisher et al., 1997). At first those areas were seemed to be related to word and face recognition alone, however there are evidences that argue that they are involved in the general process of chunking of visual objects or segregation of complex forms and can process other inputs as well (Gauthier et al., 1999). Such localisation of mental activity may explain why Broca's area is activated during various non-speech motor activities (Pulvermuller and Fadiga, 2010). In support of this view experiments analysing structural ERPs have found that the anterior cingulate gyrus is activated during lexical search (Abdullaev and Posner, 1998). Interestingly, the same area activates during conflict resolution and executive control of attention (Fan et al., 2002; Petersen and Posner, 2012). Furthermore, ERP studies of syntactic violations by Hagoort and colleagues (1993) have discovered two electric brain signatures of syntactic processing: an early left anterior negativity (LAN) and a late positive wave with a peak at 600ms (P600). LAN is a highly automated process whereas the P600 involves attention (Hahne and Friederici, 1999). In their study Hahne and Friederici (1999) manipulated the proportion of sentences with errors to sentences with no violations. Sentences with violations appeared in a low (20%) or a high (80%) violation probability conditions. The LAN effect has been elicited by both conditions, however P600 has been registered only in low probability condition. Hence, early LAN is an automated first-pass sentence parsing mechanism working in parallel with syntactic processing. At the same time, the P600 component is a second-pass parsing mechanism that directly involves executive attention. Further evidence of syntactic violation effects using fMRI (Newman et al., 2001) has revealed that they elicit significant activation in superior frontal cortex — the area that is according to some studies involved in control of attention.

To sum up the findings described above demonstrate that the brain localises processes or mental operations but not individual linguistic or non-linguistic representations. Thus, it is possible that domain-specific and domain-general mechanisms which require activation in the same area may not only share processing regions but may as well share resources between them.

Results of developmental research provide additional support to the idea that attention and language are tightly connected (Matthews and Krajewski, 2019). For example, caretakers use attentional amplification of visual input during the early stages of language development in infants. When interacting with an infant caretakers consistently direct attention to objects and/or events and follow this by corresponding names and structures. Attentional focus has been shown to improve infants' associations between objects and related words, and according to a number of studies indicates learning processes like names to objects matching (Dominey and Dodane, 2004; Estigarribia and Clark, 2007). These early foundations of the attention-language interface are later developed into a more complex linking system. The purpose of this system is to map event semantics onto sentence structure. The fundamentals of this system are being set by 2-3 years of age. Research suggests that children habitually scan referents of visually perceived transitive events following the way they are pronounced in auditory perceived utterances (Arunachalam and Waxman, 2010; Yuan and Fisher, 2009). According to Ibbotson and colleagues (2013) children as young as 3-4 years can successfully represent the perceived details of the scene via sentence structure. It can be concluded based on the role of attention in language development that there exists a strong link between the distribution of attention over the environment and the consecutive organisation of the language about this perceived environment. This link between directing attention at objects and interacting with them remains in adults - people look at objects during acting upon them irrespective of whether they explicitly speak about their actions or not (Ballard et al., 1997). Work described so far suggests that linguistic processing is following and underlying behaviour. Thus, it seems possible to draw parallels between attention mechanisms and sentence syntax.

A typical experiment on attention system implements a cueing paradigm (Posner, 1980). A cue (or a prime) in this context is something that regulates the salience of stimuli. It can be a sign (e.g., an arrow) pointing at the stimulus and hence directing attention to it, or the stimulus itself might poses some feature to attract attention (e.g., a stimulus' salience property, for example size or colour). Primes can be divided into several sub-types: exogenous (external) or endogenous (internal), explicit (consciously perceived) or implicit (unconsciously perceived); a prime can result in either overt (eye-movement to the location of the cue) or covert (attention shift without eye-movements to the location of the cue) shift of attention (Posner,1980), informative or non-informative. In my thesis I was interested in how the *modality* (visual, auditory and/or motor) of the explicit non-informative cue affects the allocation of the speaker's attention and whether it influences syntactic choice. Before we proceed, it is important to address the experiments that describe how various properties of the cues affected attention.

Perceptual priming

There are various ways of how one can manipulate the speaker's attention. In a typical priming experiment, a task for a participant is to describe a depiction of some event. At the same time their attention is cued towards one of the event's objects or referents by a prime (a cue). As earlier explained primes can be of different types (explicit/implicit etc.). Here I will describe perceptual priming work. One of the first to implement this paradigm in psycholinguistics was Prof. Russell Tomlin. He developed a variant of a perceptual priming paradigm experiment known as the "Fish Film" experiment (Tomlin, 1995). In his study, participants watched a short cartoon in which two fish approached each other and one always ate the other. Participants had to verbalize what they saw. Half of the time an explicit visual modality cue (an arrow) appeared above the Agent fish (the fish that "ate") and the other half of the time above the Patient fish (the fish that "was eaten"). The results showed that in 100% of the Agent-cued trials participants generated an active voice utterance. When the cue was on the Patient, a passive voice utterance was generated more than 90% of the time. Tomlin suggested that priming the attention tend affects the probability of the assignment of the subject role to the cued Agent or the cued Patient in an English transitive sentence production which is recognised in the choice between active and passive voice. Further, Tomlin has proposed a grammatical-role hypothesis, suggesting that attention manipulation has a direct impact on a grammatical role assigned to the cued referent.

The Fish Film paradigm has been criticised (e.g., Bock et al., 2004) and so was the grammatical-role hypothesis. One of the main criticisms was the use of the same fishfilm event without filler materials over and over, which might have led to the development of the structural bias, a tendency to repeat just produced syntactic structures (Kaschak, 2007) which in turn might have affected the proportions of passive structures. Furthermore, Tomlin used an explicit visual cue (and related experimental instructions) and presented it jointly with the cue and the target. Bock and colleagues suggested that it was quite possible that participants have simply guessed the nature of the experiment. Also, we shall understand that in real life conditions the attention is affected on much smaller scale. Thus, a later study (Gleitman et al., 2007) have improved the procedure by Tomlin (1995) avoiding most of the critical points. In their study participants described various events presented in still pictures. A visual cue (a black square) was used to manipulate the speaker's attention. The cue was presented prior to the target picture in the position of the about to be presented referent. The cue was visible for only 65ms and participants have been unaware about the manipulation. This implicit cue was still effective in directing the focus of attention to one of the referents, which is clearly seen from the eye movement data analysis. Filler materials have been included, so the probability of structure repetition has been decreased. The results have largely confirmed Tomlin's findings. However, the effects were much weaker because of the implicitness of the manipulation. On average speakers produced 10% passive voice sentences when the cue was on Patient location.

The perceptual priming literature reviewed in this section clearly demonstrates that the structure of English sentences in part reflects the speaker's attentional focus. That is, the referent's salience at least partially predicts the speaker's structural choice. In addition, the explicitness of the cue and the strength of association between the cue and the referent improve this perceptual priming effect. Finally, most recent findings suggest that the attentional system supporting the interface between referential salience and structural choice may be relatively universal and not limited to visual world events and, correspondingly, to visual attention. However, all these data have little to say about the possible mechanisms of attention-structure interplay; they only indicate that attention matters during structural selection. Gleitman et al. (2007) suggests that a hybrid system exists where perceptual priming affects both grammatical-role assignment and positioning of a constituent in a sentence. Further I will describe a number of flexible word order studies.

Perceptual priming in flexible word order languages

Research discussed above used English language predominantly. English is an SVO language with a constrained word order. During sentence production English language speaker usually chooses between active voice and passive voice SVO options and because of that the sentential starting point in many cases coincides with a subject. Because of that the hypothesis about link between attention and language (the grammatical role or the linear ordering of constituents) is not clear. However, in other languages a more flexible organisation of sentences is possible. For example, Russian, Finnish or Korean. A few studies analysed effects of perceptual priming on structural choice in Russian (Myachykov and Tomlin 2008), Finnish (Myachykov et al. 2011), and Korean (Hwang and Kaiser 2009). As English Russian and Finnish are SVO languages but both Object-initial (OVS, OSV) and Verb-initial (VSO, VOS) constructions are grammatically correct which leads to a wider range of topicalisation constructions available for a speaker. In Korean (SOV language) the verb always follows arguments. Hence, topicalisation is possible in these languages, however with some limitations. For example, discourse context factors were shown to predict ordering of sentential constituents in Korean (Jackson 2008) as well as in Finnish (Kaiser and Trueswell, 2004; Vilkuna, 1989), Russian (Comrie, 2009) and English (e.g., Downing and Noonan, 1995; Givón, 1992 etc.). Importantly, as in English speakers of Russian, Korean and Finnish can alter voice as well, but such structures are less frequent (e.g., Zemskaja, 1979; Siewierska, 1988; Vilkuna, 1989) than in English (e.g., Svartvik, 1966). Most importantly in a perceptual priming experiment in flexible word order languages it is possible for the speaker to map the cued referent both onto the subject and/or onto the starting point of the sentence without assigning the referent a subject-role. A number of relevant flexible word order languages priming studies are described below.

One of the first studies using perceptual priming paradigm in Russian is Myachykov and Tomlin (2008). In their work they proposed that if the visually cued referent is assigned a subject role, Russian speakers would similarly to English speakers alternate between active and passive voice structures when describing the Fish Film.

Another possibility was the use of topicalisation instead. This would lead increased number of Object first active voice sentences in the Patient-cued. In this scenario a linear-ordering account would be supported, suggesting that perceptual priming affects the position of the referents in the sentence, but not the syntax as such. Speakers of Russian produced 20% more Object-initial (OVS or OSV) active voice structures (with only 2% of passive voice sentences) when the cue was on the Patient. This effect is much smaller than in Tomlin (1995) where the cue led to almost 100% Passive structures, even though Myachykov and Tomlin (2008) used the same FishFilm paradigm.

In a following study Myachykov et al. (2011) compared perceptual priming effects in English and Finnish implementing the implicit priming paradigm as in Gleitman et al. (2007). English participants have shown similar pattern of results as have been reported earlier replicated earlier findings (Gleitman et al., 2007). Myachykov and colleagues (2011) have reported a main effect of Cue Location: participants produced lesser active voice sentences (74%) in the Patient-cued condition. As for Finnish speakers, a number of aspects of Finnish has to be taken into account before reporting their results. First, the passive voice in Finnish is rare and is also always produced without a structural analogue to the English by-phrase (Kaiser and Vihman, 2006). Secondly, topicalisation is equally possible in Finnish (similarly to Russian), so perceptual priming effects in Finnish may appear through topicalisation. Unfortunately, that was not the case – in Finnish although eye tracking data showed that the cueing manipulation was as effective as in English, no reliable syntax alteration happened. Similar result was observed in a study of Korean language by Hwang and Kaiser (2009). Similar implicit cueing paradigm has been implemented. Comparable results to Myachykov et al. (2009) have been achieved: implicit perceptual priming influenced oculomotor behaviour; however, this did not lead to any effect on structural choice.

In sum, the results described so far hint on the existence of differences in how visual priming affects structural choice in different languages. One possibility is that the process of coding a salient Patient referent onto the Subject role is more complicated for Russian and Finnish speakers due to the overall dispreffered usage of passive voice structures. This explanation is in line with Hwang and Kaiser (2009) who suggested that Korean has a strong bias towards SOV active voice structures and that passive structures are much more marked than in English. Thus, these studies suggest that flexible word-order languages are less influenced by the perceptual properties of the cue or referent than the fixed word-order languages.

One of the other aspects of event perception that is known to affect syntactic choice is event orientation. For example, imagine an event where a dog chases a postman: this event can have several depictions. We will focus only on the two of them, where the dog (the agent) is on the left of the postman (the patient) or on the right. The chasing event has a perceivable direction (left to right or right to left). This directionality might have an effect on the event perception and there is evidence that suggest that this factor should be accounted for, when visually mediated sentence production is studied. For example, Esaulova, Penke and Dolscheid (2019) investigated multiple factors affecting event descriptions and eye movements of the speakers. In their work one of the factors affecting the speaker was the positioning of patients in space. Thus, the position of the patient influenced eye movements, as well as time to initiate the sentence selecting the voice of the sentence. Esaulova et al (2019) concluded that participants have a bias to expect agents to the left of patients in visual scenes and to assign subject functions to left-positioned rather than right-positioned referents. Moreover, some earlier finding suggest that leftmost position in space is associated with agents rather than patients (Dobel et al, 2007). In their study participants had to draw out the heard sentences. Agents were positioned to the left of the patients most of the times. Also, further support of this view comes from experiments by Chatterjee et al. (1995, 1999). Their results hint that positioning the agent on the left speeds up the recognition and makes it less effortful for the speaker in comparison to when the agent is on the right. It is then important to at least keep in mind how this factor during stimuli development and analysis of the collected data. Also, it is interesting to investigate this issue in Russian as to our knowledge nothing like that has been investigated before.

General characteristics of dissertation

Theoretical novelty

Above studies hint on the existence of a regular interface between visual attention and syntactic choice during visually mediated sentence production. Several studies have demonstrated how speakers regularly choose between structural alternatives because of attention shift. In English, the referent is assigned to the subject role in a produced sentence more often after a cue. What is yet to find out is how this mapping from attentional domain to the linguistic domain takes place in other languages. One possibility is that different languages' grammars use different grammatical encoding mechanisms of perceptual properties of the described world (Tomlin, 1997). Another possibility is that the mechanisms are somewhat universal and speakers most of the time map the salient (or cued) referent onto subject. The difference as we have seen above may be in the availability of the alternatives (e.g., active vs. passive) across languages.

Mostly, previous research focused on effects of different types of cues on language production (Myachykov et al 2011). The cues have been delivered to participants visually via referent preview (Myachykov et al, 2012a), verb preview (Myachykov et al, 2012b), locative cues (Tomlin, 1995; Gleitman et al, 2007; Myachykov et al, 2012b). However, sentence production process might also be influenced by cues from other modalities (e.g. auditory/acoustic or motor) which in turn might further support the interaction between attention and language (Myachykov et al, 2012b). Thus, it is important to investigate if attentionally motivated structural choice is limited to the visual modality or whether the attentional system sub-serving structural choice is more general. Results of a number of recent studies corroborate the universalist general interaction system view (e.g., Spence, 2010; Kostov and Janyan, 2012). For example, in their study of manipulation affordances, Kostov and Janyan (2012) found that using a lateral auditory cue leads to the establishment of a lateral affordance effect typically observed in studies using visual input and attention manipulations. Similarly, Spence (2010) studied unimodal and bimodal sensory cues from auditory, visual, and motor modalities. Results of this study suggests that unimodal (one cue from one modality at a time), as well as bimodal (two cues from two different modalities), successfully orient attention to the source of the cue. Also, previous research studied how cues from different attention modalities interact with each other (Driver and Spence, 2004; Fritz et al., 2007; Shinn-Cunningham, 2008).

Hypothesis

The present project aimed at examining if and how auditory and motor cues affect syntactic choice in comparison to visual cues. In the series of perceptual priming experiments English and Russian native speakers described transitive events. As primes visual, auditory or motor cues in English language studies (Experiments 1, 2 and 3) and unimodal cues (visual, auditory or motor (Experiments 4, 5 and 6 respectively) in Russian was used to direct attention to referents. As it was discussed above according to the universalist view of attention and sensory integration attention can shift to the location primed by visual, auditory, and motor cues. As such, this project focused on whether syntactic structure of a sentence can be biased shifts of attention caused by auditory and/or motor attentional cues, as well as whether a combination of cues presented simultaneously might affect the syntactic choice differently. This latter focus motivated the choice of specific cueing manipulations. I was interested in how and if attentional cues of motor and auditory modalities bias attention and, as a result, syntactic choice. Also, we wanted to investigate if bimodal cueing is more effective than unimodal. Last but not the least, we have controlled our stimuli for left to right and right to left event orientation and investigated the role of this factor in Russian language.

Method

In English language studies (Experiments 1, 2 and 3) two factors were manipulated: Cue Type (Exp.1 visual and auditory, Exp.2 auditory and motor, Exp.3 visual and motor modality) and Cue Location (cues were presented either congruently or incongruently). Hence a 2x2 factorial design with type of the cue and location of the cue as both within-subjects and within-items factors was used. For the dependent variable in English language studies we chose passive voice structures similarly to previous experiments (i.e. Tomlin, 1995).

For Russian language experiments (Experiments 4, 5, and 6) we have adopted the same procedure as in English experiments, with one main difference that only one type of cues has been used per experiment, as before testing the combined cues it was important to test if in Russian language single modality cues will be efficient in affecting produced sentences. Two independent variables were manipulated: Cue Location (cue on agent vs. cue on patient) and Event Orientation (Agent left vs. Agent right). Event orientation has been used as a factor in Experiments 4, 5 and 6 because according to Tversky (2011) the event orientation might be an important contributor to syntactic choice, as there exists a canonical event perception direction, based on the reading direction in languages. Again, as in English language experiments a 2x2 factorial design have been chosen. This time, however, due to peculiarities of Russian language (complicated passivisation, free word order) the DV was the proportion of the sentences where Object was named first in the sentence (Patient-initial sentences).

Participants

This project has been approved by Northumbria University and Higher School of Economics ethics committees. Overall, we have recruited 144 participants native English (N=72) and Russian (N=72) speakers from the population of students and staff of Northumbria University and Higher School of Economics took part in 6 experiments. Some exclusion criteria have been applied. Firstly, participants had to be monolingual native speakers of either Russian or English. Secondly, they had to have normal (or corrected to normal) vision. We have included participants with glasses and contact lenses, however not all of them could pass the eye tracking calibration procedure, so their data has been excluded from later analysis. Finally, participants had to have no history of language or attention-related disorders (e.g., dyslexia, ADHD). In all but Experiment 1, participants received a monetary reward (£5 or 125 roubles). In Experiment 1 participants were given participation points. Written informed consent has been obtained from all participants.

Materials

All materials including depiction of events, characters, event names, and filler pictures have been provided by Myachykov and colleagues from their similar work (2012). Out of the provided materials six types of transitive events were picked excluding ditransitive events (chase, touch, kick, shoot, push, hit). Then the list of sixteen referents (nun, clown, pirate, artist, cowboy, monk, policeman, swimmer, dancer, professor, waitress, boxer, chef, soldier, burglar) from the provided material have been picked. Those 16 referents have been paired and assigned to 6 events, creating 48 stimuli (for an example stimuli picture see Figure 1). Event orientation was equally balanced between left-to-right and right-to-left depictions. Moreover, following Bock and colleagues' critique (2004) of Tomlins (1995) work we have included 98 filler pictures. Two filler pictures have always preceded the target pictures. The order of target pictures was

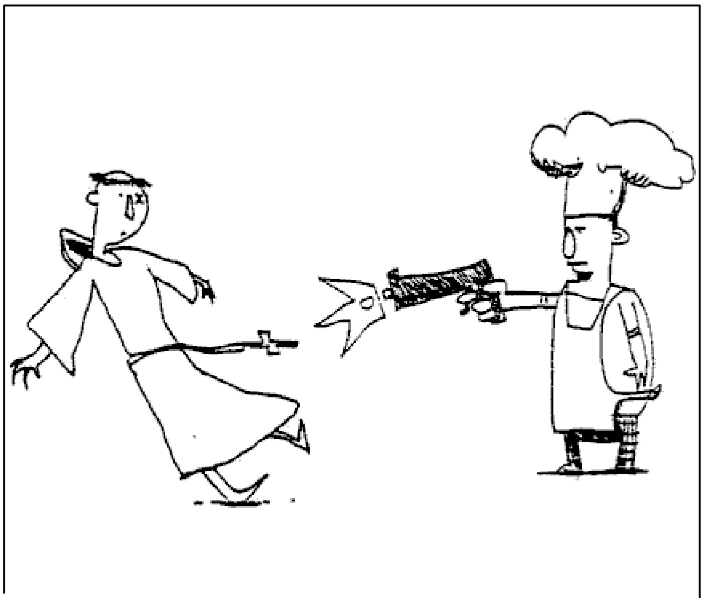

Fig. 1. Stimuli: "The chef shoots the monk".

randomised. Each stimuli picture had to be in all four experimental conditions for a sensible analysis. However, participants were not supposed to see the stimuli twice, so each picture were assigned to one of the 4 lists. Those lists have been developed using a latin-square design – each event was shown to the participant only once in one of the four conditions. Each participant saw only one list.

Apparatus

Experiments were conducted on the premises of the CoCo lab in Northumbeia university or in the eye tracking lab of the CDM centre in the National Research University Higher School of Economics. All experiments but Experiment 1 has been created in SR Research Experiment Builder v1.8 software (SR Research Ltd, Ottawa, Canada). Experiment 1 has been developed in Microsoft Power Point (Microsoft, Inc.). An EyeLink 1000 Desktop eye-tracker (at Northumbria) or EyeLink 1000+ (in HSE) (SR Research Ltd, Ottawa, Canada) were used to record the participant's gaze at 1MHz sampling rate. Throughout experiments we recorded the right eye of the participants. The calibration error threshold has been set to 1 degree of visual angle, however on average it never exceeded 0.7 degree of visual angle. All eye tracking data filters were set at default mode.

To present the stimuli we used Clear view 17-inch display with a refresh rate of 60Hz (in Northumbria) or a 24-inch 144hz ASUS VG248 (in HSE). Participants were seated on an adjusted seat, with their head in the chin rest, 60cm eye-to-monitor distance. For Auditory stimuli presentation we used Genius E120 Desktop Speakers (in Northumbria) or Sven 312 speakers (in HSE). The speakers (where applicable) were positioned below the monitors 15cm to the left and to the right from the centre of the screen. This position was dictated by the setups of auditory cueing studies (Santangelo and Spence, 2007; Spence and Santangelo, 2009; Spence, 2010) in order to approximate the auditory cue as close as possible to the cued referent. Before the experiment, the sound level was adjusted for the participant to clearly hear auditory cues. The sound was processed by ASIO sound card. Responses were recorded using a voice recorder software (Voice Record PRO, various versions VENDOR) installed on a Samsung smartphone.

For motor responses participants used either the keyboard or the xbox 360 controller. The collected data was stored on a password protected server within the premises of HSE for according to the data protection regulations. No sensitive personal information has been recorded, collected, or stored.

Procedure

The procedure was the same in Russian and English experiments. On arrival, participants provided their demographic information (age and gender) and signed consent forms. Participants then read experimental instructions, then went through practice, calibration, and validation procedure (5-10 minutes depending on the ease of calibration). Throughout the practice session, participants were shown examples of referents, events and stimuli and practiced their responses. The referent's depictions were sequentially presented on the monitor. Participants read the names which were written below the picture aloud and to try and remember them, to use later in the experiment. Then, participants practiced the names of the events. Participants saw fourteen randomly presented events (filler events included) with their names (e.g., chase) written below. Upon completion of the practice session, participants were given experimental instructions. Depending on which experiment they participated in, that they would be presented with either a visual cue (VC, red circle 100x100 pixels, presented on the relative positions of the referents) an auditory cue (AC, a "beep" sound delivered from the left or from the right speaker) or a motor cue (MC, coloured circle requiring a button press) or a combination of the above. The participants were instructed to avoid detailed descriptions in their responses (i.e. they were instructed to produce simple descriptive sentences like "The policeman kicks the robber" instead of "The policeman with a club in black coat kicks the robber, and the robber seems to be confused"). Finally, participants were presented with several training trials. This was done for participants to familiarise with the procedure (see Figure 2 for a visualisation of a trial) and to train their responses. Below, experiments are described in a more detailed fashion.

In visual perceptual priming experiments participants were instructed that they will see black dot in the middle of the screen (fixation circle, fixation confirmed by the fixation trigger via the eye tracker, gaze contingent offset after 150ms fixation maintenance), followed by a red circle to the left or to the right of central fixation. They were instructed to look at the circle as soon as they saw it and to wait for the picture.

In auditory experiments an auditory signal (a tone of 500ms duration, sample rate 22050Hz) was presented from the left or from the right speaker. As before, participants were instructed to look to the black dot (fixation circle) in the middle of the screen and wait. As soon as fixation triggered (gaze contingent offset after 150ms fixation), the lateral auditory cue sounded. Participants were informed about the cue coming from either the left or the right speaker and were instructed to locate the sound.

In motor cueing experiments participants pressed the green or the blue key (Z and NumPad2 on the keyboard or L1 or R1 buttons on XBOX controller in Experiments 3 and 6) depending on the colour of a circle which appeared in the centre of the screen after the fixation dot. The 'Z' (L1) key was used in the left cue condition and the 'NumPad2' (R1) key for the rightward cue condition. The buttons were to be pressed with corresponding index fingers. In Experiments 3 and 6 (English language Visual vs. Motor modality priming and Russian language motor priming experiments) the keyboard has been replaced with the XBOX 360 joystick and participants have been prompted to press

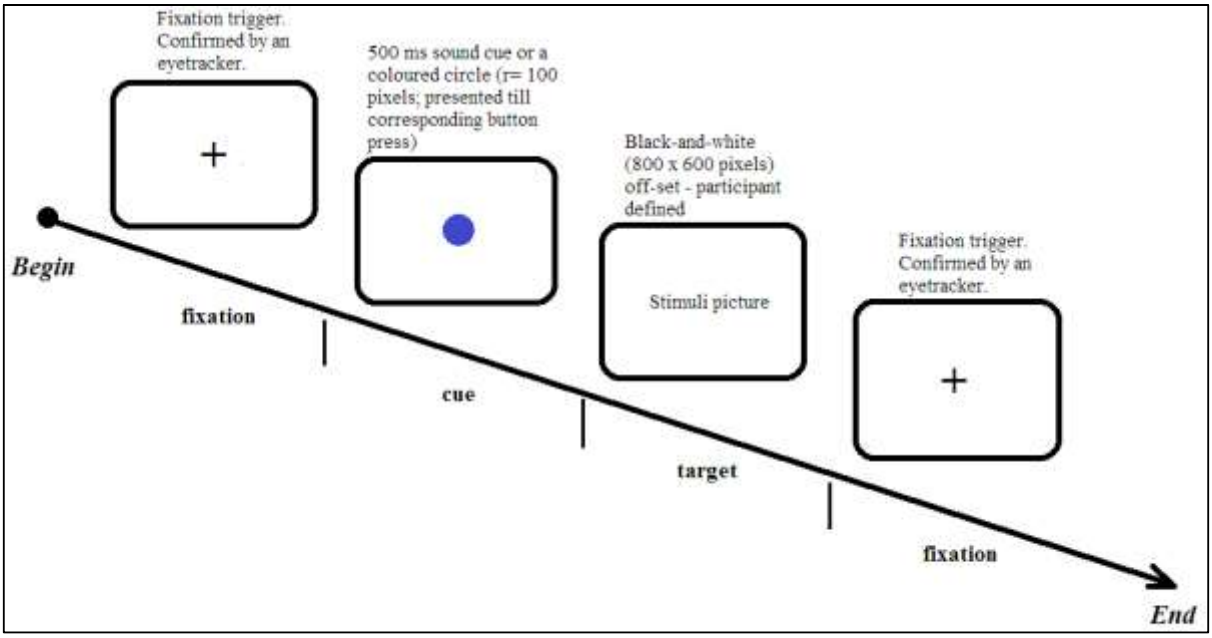

Fig. 2. A diagram of the experimental procedure.

either left or right trigger buttons (L1 and R1). This has been done to further reduce the probability of participants gazes at the keyboard, as the joystick is handheld.

In all experiments the priming part was followed by the target event presentation. Participants have been instructed to describe the event as soon as they saw it with one sentence. Participants proceeded to the next trial as soon as they finished by pressing Space or an X button on the xbox 360 controller.

Analysis

The recorded responses were thoroughly processed and coded as either Active Voice or Passive Voice (Experiments 1, 2 and 3) or Agent-initial or Patient-initial (Experiments 4, 5 and 6). Active Voice descriptions were coded as such if the transcribed description included a subject (agent) noun phrase, followed by a transitive verb, and an object (patient) noun phrase. Passive Voice descriptions were coded as such if they included a subject (patient) noun phrase, followed by a passivised transitive verb, and a by-phrase including the agent. Russian language responses were coded as Agent-initial sentence if the agent referent have been named before the patient (e.g. SVO "Полицейский бьет вора" (Policeman kicks robber), SOV "Полицейский вора бьет" (Policeman robber kicks) structures). If patient has been named before agent, the response was coded as Patient-initial. Conjoined noun phrases (e.g. The policeman and the robber are fighting), truncated passives (e.g. The robber is being kicked), sentences with grammar mistakes, and responses consisting of several sentences were. Those are accounted for less than 1% of all recorded responses.

All experiments have been analysed using Generalised Linear Mixed Effects Models (GLMM), as part of the lme4 package in R. All analysis procedures were conducted in RStudio software (version 1.2.1335). The DV was the occurrence of a Passive Voice or Patient-initial description (True = 1, False = 0). Thus, among available models in the glmer() function a binary logistic model was chosen. The model included a full-factorial Cue Modality (Visual, Auditory, Motor) × Cue Location (Agent, Patient) and Event Orientation (Left to right x Right to left) in Russian language experiments fixed effects design. All predictors were mean-centred using deviation-coding. The maximal random effects structure (Barr et al., 2013) was used. All experimental variables were within-subjects and within-items. Thus the model included by-subject and by-item random intercepts, by-subject and by-item random slopes for every main effect and interaction term in the fixed effects design. Random correlations were also included. Likelihood Ratio Chi-Square (${}_{LR}\chi 2$) model comparisons have provided the p-values.

Results

Figures 3-5 show the results of the English language experiments. Graphs represent the proportions of produced passive voice to active voice syntactic structures in percents. In Experiment 1 (Fig.3), the effect of bimodal cueing of attention was evaluated. Subjects were primed with auditory and visual cues either in congruent (both markers on the same referent) or incongruent conditions. Descriptive statistics show that the syntactic choice have been affected only when the focus of attention has been shifted by means of a visual cue. In Experiment 2 attention was primed by motor and auditory cues. Once again, we can see on the Fig.4 that the auditory marker did not affect the grammatical structure of the sentence. Experiment 3 was evaluating how visual and motor cues interact. The position of the cues affected the choice of grammatical structure, however, only when visual cue was priming the patient referent (Fig.5).

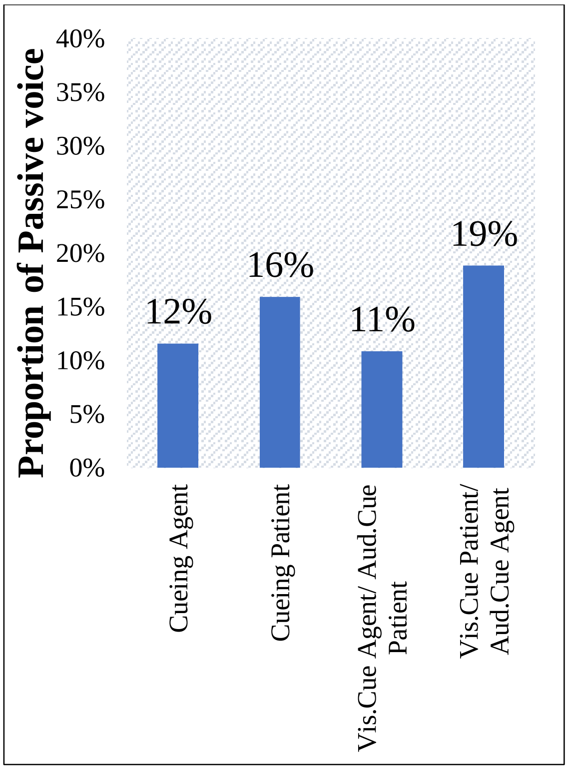

Fig. 3. Auditory and visual priming Experiment 1 (Article 2).

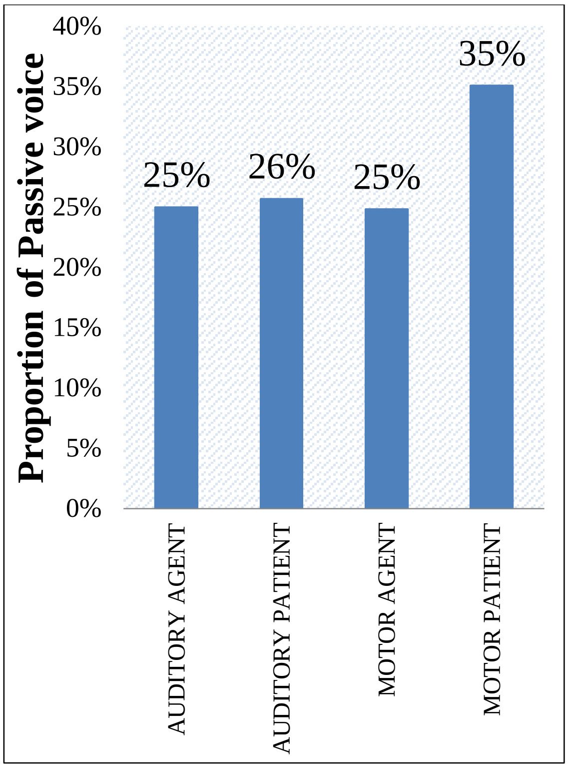

Fig. 4. Auditory and Motor priming Experiment 2 (Article 1).

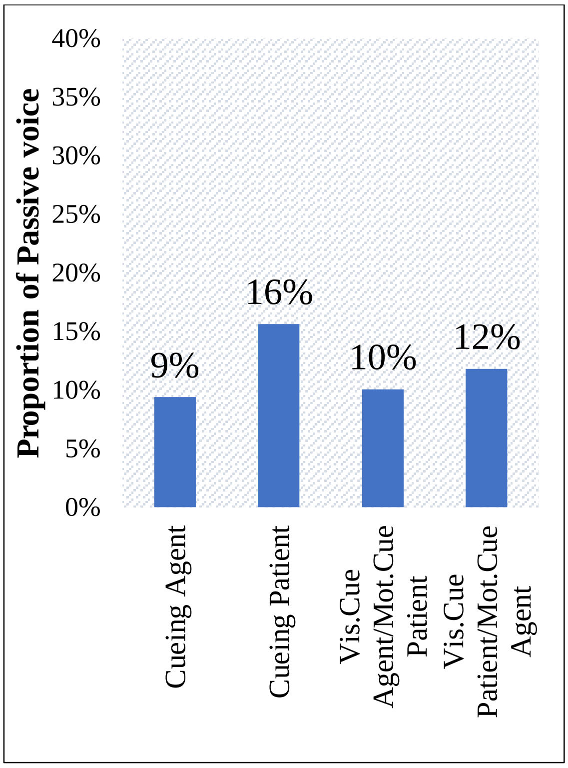

Fig. 5. Visual and Motor priming Experiment 3 (article in prep.).

The results of statistical analysis of English language studies are presented in Table 1. Analysis of all experiments was carried out in R studio software. All data was subjected to linear modeling using the lme4 package. The DV of was the occurrence of a Passive Voice description. It was coded 1 in case when it occurred and 0 for all other cases. Since the results of the dependent variable were encoded as the absence (0) or the presence (1) of the required syntactic structure, a binomial linear model was chosen. In the table statistically significant effects are highlighted in bold.

Tab. 1. Results of inferential GLMM lme4.

Language	Exp №	Priming modality	IV	GLMM lme4
English	1	Visual and auditory	**Visual cue (Ag/Pat)**	**$*\chi2(1) = 4.31$, p=.04**
			Auditory cue (Ag/ Pat)	$\chi2(1) = .306$, p=.58
			Interaction	$\chi2(1) = .068$, p=.79
	2	Motor and auditory	**Cue location (Ag/Pat)**	**$*\chi2(1) = 5.29$, p=.02**
			Cue type (Aud/Mot)	$\chi2(1) = 1.401$, p=.23
			Interaction	**$*\chi2(1) = 4.940$, p=.02**
	3	Visual and motor	**Visual cue (Ag/Pat)**	**$*\chi2(1) = 4.045$, p=.04**
			Motor cue (Ag/Pat)	$\chi2(1) = 1.241$, p=.26
			Interaction	$\chi2(1) = .473$, p=.49

Figures 6 to 8 show the results of Russian language experiments. The graphs show the proportions of the patient-initial sentences. Independent variables the cued referent (i.e. Cue Location), and the direction of the event (Left-to-Right or Right-to-Left). In Experiment 4, the effects of visual cueing of attention and event orientation on the order of mentioning of referents in the generated sentence were evaluated. Descriptive statistics (Fig.6) confirmed the effect cue location on syntactic choice - the highest number of patient-initial sentences was generated when the cue was on the patient (35% versus 19%). You can also trace the effect of the event-orientation on syntactic choice. In Experiment 5, an auditory cue was used. Please note, similarly to English language experiment, the auditory cue did not affect the choice of grammatical structure (Fig.7) - the difference between the position of the marker and the sentences generated is 1%. However, it is worth highlighting the effect of the event orientation on the grammatical structure, like Experiment 4 (19% versus 35% of the patient initial sentences). Experiment 6 tested the effect of a motor priming on grammatical structure in Russian language. An interesting result is that, similarly to English language motor experiments (Experiments 2 and 3), the motor cue influenced the grammatical structure, but the effect of the event orientation has completely disappeared (18% versus 19%).

Figures 6-8. Group-average probability of patient-initial structures as a function of Cue Location and Event orientation. Russian language experiments.

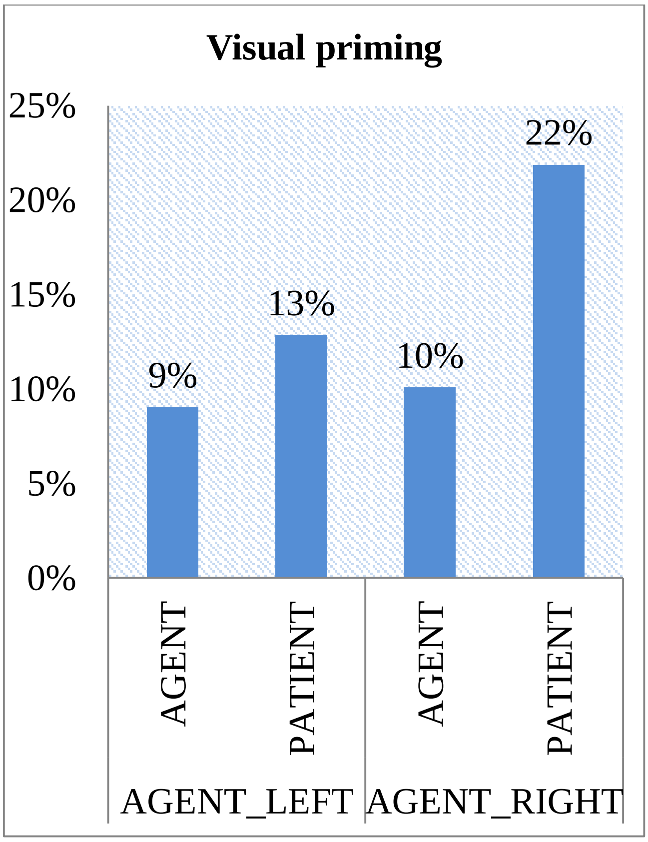

Fig. 6. Visual priming Experiment 4 (Article 3).

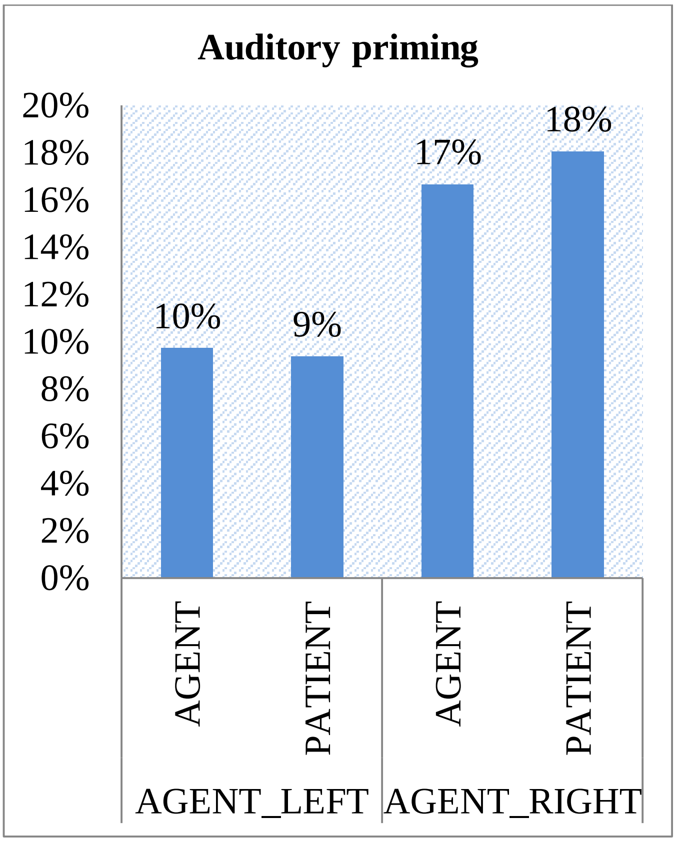

Fig. 7. Auditory priming Experiment 5 (article in prep.).

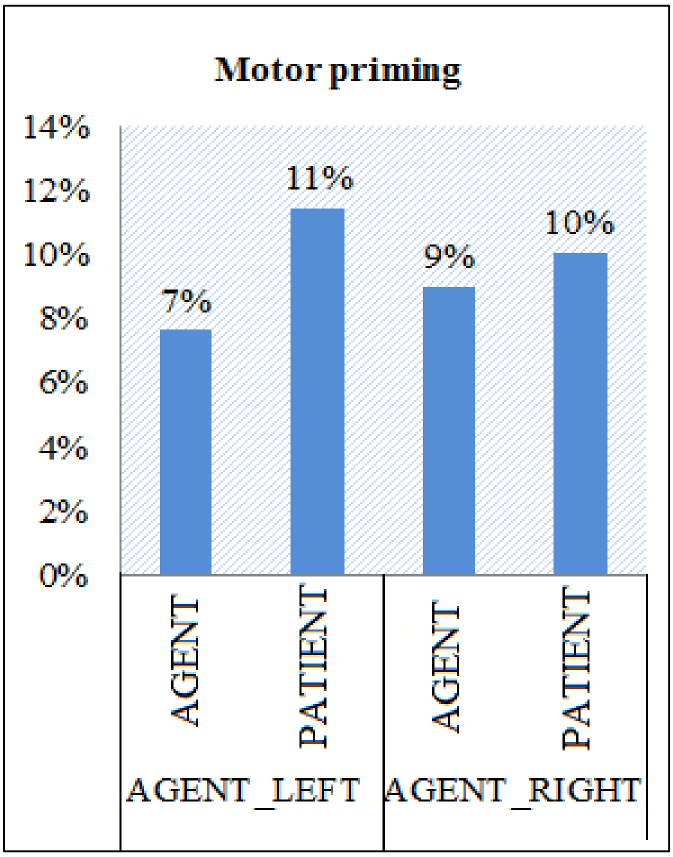

Fig. 8. Motor priming Experiment 6 (article in prep.).

The results of statistical analysis of English language studies are presented in Table 1. Analysis of all experiments was carried out in R studio software. All data was subjected to linear modeling using the lme4 package. The DV of was the occurrence of a Patient initial description. It was coded 1 in case when it occurred and 0 for all other cases. Since the results of the dependent variable were encoded as the absence (0) or the presence (1) of the required syntactic structure, a binomial linear model was chosen. In the table statistically significant effects are highlighted in bold.

Table 2. Results of inferential GLMM lme4.

Language	Exp №	Priming modality	IV	GLMM lme4
Russian	1	Visual	**Cue location (Ag/Pat)**	***$\chi2(1) = 13.051, p<.001$**
			Event orientation	***$\chi2(1) = 4.404, p=.04$**
			Interaction	$\chi2(1) = 2.056, p=.15$
	2	Auditory	Cue location (Ag/Pat)	$\chi2(1) = .099, p=.75$
			Event orientation	***$\chi2(1) = 6.86, p=.01$**
			Interaction	$\chi2(1) = .146, p=.703$
	3	Motor	**Cue location (Ag/Pat)**	***$\chi2(1) = 4.04, p=.04$**
			Event orientation	$\chi2(1) = 0.1, p=.99$
			Interaction	$\chi2(1) = 1.441, p=.20$

Key conclusions and outcomes

In this thesis, I tried to answer several questions about the influence of the modality of attention, as well as other variables on the syntactic choice in Russian and English. As a result of this work, several important conclusions can be done. First, perceptual priming by means of visual modality cues affects syntactic choice in both English and Russian languages. Participants in my experiments produced more passive voice sentences and patient first sentences after the presentation of the visual cue on the location of the patient, independent of structural bias and event orientation in Russian. These results replicate the earlier findings in English and Russian languages, (Myachykov and Tomlin, 2008; Myachykov et al., 2012 a,b; Tomlin, 1995; Gleitman et al., 2007) which supports the grammatical role hypothesis of Tomlin but not the linear ordering hypothesis.

Auditory priming of attention does not affect syntactic choice both in Russian and in English. A possible explanation is that the auditory cueing is not strong enough to guide attention towards the cued referent. Another possible explanation is the possible experimental flaws. Due to the duration of the cue (500ms) it is quite possible that participant's attention has been drawn away from the cued location and inhibition of return prevented the cued referent to be processed first. Also, several participants in the post experiment conversation said that they stopped "hearing" the cue after some time in the experiment. This can be an interesting issue to address in further attention manipulation research. It seems that participant's attention have inhibited the auditory prime due to some physical aspects of the stimuli or due to its "unrelatedness" (according to some participants, who said they thought the sound was to distract them) to the task. Hence, a more implicit auditory cue may result in a successful attention shift which in turn may be reflected by the sentence structure.

According to the results of motor priming experiments – shifts of attention caused by motor responses affect syntactic choice. Subjects generated more passive (in English) and more patient-initial (in Russian) sentences after corresponding primes.

The study of bimodal priming in English language has not yielded significant results. Only in a visual-motor experiment have we registered an effect in a congruent condition: provided that the visual and motor cues had been focusing attention on the patient, the largest number of passive voice sentences were generated in comparison with other conditions. This suggests that perhaps the assignment of the subject role is affected by the strongest cue provided at a given moment and in the case of rivalry between different modalities, a weaker cue is suppressed.

The direction of the event is an important factor in choosing the syntactic structure in Russian language. If you look at the results of the experiment with auditory priming, the cue did not provide an effect, but the direction of the event affected the choice of grammatical structure. This goes in concordance with previous studies (e.g. Esaulova et al, 2019) in other languages and thus hints on the universality of this effect in language.

Statements for defense

To sum up it is important to notice that by this thesis I do not postulate the existence of different attentional systems e.g. motor, auditory etc. I am rather suggesting that various modality inputs feed information to the singular attention mechanism, which due to the limitation of processing capacities must be peaky in order to provide the most relevant information to the language production system. Hence, from my results it is quite clearly seen. It can be suggested that for visually mediated sentence production *visual* information is of most importance. When auditory input is received it is inhibited by the system as a distractor because of its lack of informativeness, even though participants have been informed during the training procedure about the cue coming either from the left or from the right speaker., and thus providing information about cued referent location. Motor prime on the other hand when participants used the keyboard to respond had a tighter connection to the visual system, as participants quite possibly attended to the keyboard in order to press the correct button by looking at it which created a stronger link between the motor prime and the referent. Moreover, there might have been an effect of colour on the response. The visible light (colour) spectrum goes left to right - from blue, to green, to red. Primes in my motor experiment were blue and green, which are located to the left (blue) and in the middle (green) of the light (colour) spectrum. It is possible that some embodied effects have been evoked by different colours which is an interesting idea to address in future analysis and research.

What we can clearly conclude from this research is that attention and language are directly related, and their relation can be seen in the produced speech. Further investigation will look deeper in the specific mechanisms of this relation. Which system of attention is responsible for this relation? Our idea is that by using attention network test (Fan et al., 2002) and correlating results on it with produced responses we may see a correlation between specific attention network and for example the amount of produced patient initial sentences. For example, a participant showing faster reaction times on executive control network task will produce more passive voice structures. This will hint us on looking deeper into the link between executive control network and sentence production mechanism by using magnetic or electrical stimulation/inhibition methods to find out how exactly attention and language are related in the brain.

References

Abdullaev Y. G., Posner M. I. Event-related brain potential imaging of semantic encoding during processing single words //Neuroimage. – 1998. – Т. 7. – №. 1. – С. 1-13.

Arunachalam S., Waxman S. R. Meaning from syntax: Evidence from 2-year-olds //Cognition. – 2010. – Т. 114. – №. 3. – С. 442-446.

Ballard D. H. et al. Deictic codes for the embodiment of cognition //Behavioral and brain sciences. – 1997. – Т. 20. – №. 4. – С. 723-742.

Barr D. J. et al. Random effects structure for confirmatory hypothesis testing: Keep it maximal //Journal of memory and language. – 2013. – Т. 68. – №. 3. – С. 255-278.

Bock K., Irwin D. E., Davidson D. J. Putting first things first. – 2004.

Comrie B. (ed.). The world's major languages. – Routledge, 2009.

Chatterjee A., Maher L. M., Heilman K. M. Spatial characteristics of thematic role representation //Neuropsychologia. – 1995. – Т. 33. – №. 5. – С. 643-648.

Chatterjee A., Southwood M. H., Basilico D. Verbs, events and spatial representations //Neuropsychologia. – 1999. – Т. 37. – №. 4. – С. 395-402.

Dobel C., Diesendruck G., Bölte J. How writing system and age influence spatial representations of actions: A developmental, cross-linguistic study //Psychological Science. – 2007. – Т. 18. – №. 6. – С. 487-491.

Dominey P. F., Dodane C. Indeterminacy in language acquisition: the role of child directed speech and joint attention //Journal of Neurolinguistics. – 2004. – Т. 17. – №. 2-3. – С. 121-145.

Downing P. A., Noonan M. (ed.). Word order in discourse. – John Benjamins Publishing, 1995. – Т. 30.

Esaulova Y., Penke M., Verlage S. Describing events: changes in eye movements and language production due to visual and conceptual properties of scenes //Frontiers in psychology. – 2019. – Т. 10. – С. 835.

Estigarribia B., Clark E. V. Getting and maintaining attention in talk to young children //Journal of child language. – 2007. – Т. 34. – №. 4. – С. 799-814.

Fan J. et al. Testing the efficiency and independence of attentional networks //Journal of cognitive neuroscience. – 2002. – Т. 14. – №. 3. – С. 340-347.

Fritz J. B. et al. Auditory attention—focusing the searchlight on sound //Current opinion in neurobiology. – 2007. – Т. 17. – №. 4. – С. 437-455.

Gauthier I. et al. Activation of the middle fusiform'face area'increases with expertise in recognizing novel objects //Nature neuroscience. – 1999. – Т. 2. – №. 6. – С. 568.

Givón T. The grammar of referential coherence as mental processing instructions //Linguistics. – 1992. – Т. 30. – №. 1. – С. 5-56.

Gleitman L. R. et al. On the give and take between event apprehension and utterance formulation //Journal of memory and language. – 2007. – Т. 57. – №. 4. – С. 544-569.

Hagoort P., Brown C., Groothusen J. The syntactic positive shift (SPS) as an ERP measure of syntactic processing //Language and cognitive processes. – 1993. – Т. 8. – №. 4. – С. 439-483.

Hahne A., Friederici A. D. Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes //Journal of cognitive neuroscience. – 1999. – Т. 11. – №. 2. – С. 194-205.

Hwang H., Kaiser E. Accessibility effects on production vary cross-linguistically: Evidence from English and Korean //Journal of Memory and Language. – 2015. – Т. 84. – С. 190-204.

Ibbotson P., Lieven E. V. M., Tomasello M. The attention-grammar interface: eye-gaze cues structural choice in children and adults. – 2013.

Jackendoff R. The architecture of the linguistic-spatial interface //Language and space. – 1996. – Т. 1. – С. 30.

Jackson K. H. The effect of information structure on Korean scrambling. – University of Hawaii at Manoa, 2008.

Kaiser E., Vihman V. A. Invisible arguments: Effects of demotion in Estonian and Finnish //Demoting the agent. – John Benjamins, 2006. – С. 111-141.

Kaiser E., Trueswell J. C. The role of discourse context in the processing of a flexible word-order language //Cognition. – 2004. – Т. 94. – №. 2. – С. 113-147.

Kanwisher N., McDermott J., Chun M. M. The fusiform face area: a module in human extrastriate cortex specialized for face perception //Journal of neuroscience. – 1997. – Т. 17. – №. 11. – С. 4302-4311.

Kaschak M. P. Long-term structural priming affects subsequent patterns of language production //Memory & Cognition. – 2007. – Т. 35. – №. 5. – С. 925-937.

Kostov K., Janyan A. The role of attention in the affordance effect: can we afford to ignore it? //Cognitive processing. – 2012. – Т. 13. – №. 1. – С. 215-218.

Landau B., Jackendoff R. “What” and “where” in spatial language and spatial cognition //Behavioral and brain sciences. – 1993. – Т. 16. – №. 2. – С. 217-238.

Levelt W. J. M. Speaking: From intention to articulation. – MIT press, 1993. – Т. 1.

Levelt W. J. M. Speaking: From intention to articulation. ACL. – 1989.

Matthews D., Krajewski G. First language acquisition //Cognitive Linguistics-A Survey of Linguistic Subfields. – 2019. – С. 159.

McCandliss B. D., Cohen L., Dehaene S. The visual word form area: expertise for reading in the fusiform gyrus //Trends in cognitive sciences. – 2003. – Т. 7. – №. 7. – С. 293-299.

Myachykov A. et al. Visual attention and structural choice in sentence production across languages //Language and Linguistics Compass. – 2011. – Т. 5. – №. 2. – С. 95-107.

Myachykov A., Garrod S., Scheepers C. Determinants of structural choice in visually situated sentence production //Acta psychologica. – 2012a. – T. 141. – №. 3. – C. 304-315.

Myachykov A. et al. Referential and visual cues to structural choice in visually situated sentence production //Frontiers in psychology. – 2012b. – T. 2. – C. 396.

Myachykov A., Tomlin R. S. Perceptual priming and structural choice in Russian sentence production //Journal of Cognitive Science. – 2008. – T. 6. – №. 1. – C. 31-48.

Myachykov A., Garrod S., Scheepers C. Attention and memory play different roles in syntactic choice during sentence production //Discourse Processes. – 2018. – T. 55. – №. 2. – C. 218-229.

Myachykov A., Pokhoday M., Tomlin R. Attention and structural choice in sentence production //The Oxford Handbook of Psycholinguistics. – 2018.

Myachykov A., Posner M. I., Tomlin R. S. A parallel interface for language and cognition in sentence production: Theory, method, and experimental evidence //The Linguistic Review. – 2007. – T. 24. – №. 4. – C. 457-474.

Myachykov A. et al. Syntactic flexibility and competition in sentence production: The case of English and Russian //The quarterly journal of experimental psychology. – 2013. – T. 66. – №. 8. – C. 1601-1619.

Myachykov, A. Integrating perceptual, semantic and syntactic information in sentence production. Diss. University of Glasgow. – 2007.

Myachykov A., Garrod S., Scheepers C. Perceptual priming of structural choice during English and Finnish sentence production //Language & cognition: State of the art. – 2010. – C. 54-72.

Newman A. J. et al. An event-related fMRI study of syntactic and semantic violations //Journal of psycholinguistic research. – 2001. – T. 30. – №. 3. – C. 339-364.

Petersen S. E., Posner M. I. The attention system of the human brain: 20 years after //Annual review of neuroscience. – 2012. – Т. 35. – С. 73-89.

Posner M. I. Orienting of attention //Quarterly journal of experimental psychology. – 1980. – Т. 32. – №. 1. – С. 3-25.

Pulvermüller F., Fadiga L. Active perception: sensorimotor circuits as a cortical basis for language //Nature Reviews Neuroscience. – 2010. – Т. 11. – №. 5. – С. 351.

Santangelo V., Spence C. Multisensory cues capture spatial attention regardless of perceptual load //Journal of Experimental Psychology: Human Perception and Performance. – 2007. – Т. 33. – №. 6. – С. 1311.

Shinn-Cunningham B. G. Object-based auditory and visual attention //Trends in cognitive sciences. – 2008. – Т. 12. – №. 5. – С. 182-186.

Siewierska A. The passive in Slavic //Passive and voice. – 1988. – Т. 243. – С. 89.

Spence C., Santangelo V. Capturing spatial attention with multisensory cues: A review //Hearing research. – 2009. – Т. 258. – №. 1-2. – С. 134-142.

Spence C. Crossmodal spatial attention //Annals of the New York Academy of Sciences. – 2010. – Т. 1191. – №. 1. – С. 182-200.

Spence C. J., Driver J. Covert spatial orienting in audition: Exogenous and endogenous mechanisms //Journal of Experimental Psychology: Human Perception and Performance. – 1994. – Т. 20. – №. 3. – С. 555.

Svartvik J. On voice in the English verb. – Walter de Gruyter, 2011. – Т. 63.

Tomlin R. S. Focal attention, voice, and word order: an experimental, cross-linguistic study //Word order in discourse. – 1995. – Т. 30. – С. 517-554.

Tomlin R. S. Mapping conceptual representations into linguistic representations: The role of attention in grammar //Language and conceptualization. – 1997. – С. 162-189.

Tversky B. Visualizing thought. Topics in Cognitive Science, 3 (3), 499-535. – 2011.

Vilkuna M. Free word order in Finnish. – 1989.

Yuan S., Fisher C. “Really? She blicked the baby?” Two-year-olds learn combinatorial facts about verbs by listening //Psychological science. – 2009. – Т. 20. – №. 5. – С. 619-626.

Zemskaja E. A. Russkaja razgovornaja reč: lingvističeskij analiz i problemy obučenija. – Russkij jazyk, 1979.