



NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

*Mikhail Pokhoday, Christoph Scheepers
Yury Shtyrov, Andriy Myachykov*

MOTOR (BUT NOT AUDITORY) ATTENTION AFFECTS SYNTACTIC CHOICE

BASIC RESEARCH PROGRAM

WORKING PAPERS

**SERIES: PSYCHOLOGY
WP BRP 80/PSY/2017**

This Working Paper is an output of a research project implemented at the National Research University Higher School of Economics (HSE). Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE

*Mikhail Pokhoday¹, Christoph Scheepers²,
Yury Shtyrov³, Andriy Myachykov⁴*

MOTOR (BUT NOT AUDITORY) ATTENTION AFFECTS SYNTACTIC CHOICE⁵

Understanding the determinants of syntactic choice in sentence production is a salient topic in psycholinguistics. Existing evidence suggests that syntactic choice results from an interplay between linguistic and non-linguistic factors, and a speaker's attention to the elements of a described event represents one such factor. Whereas multimodal accounts of attention suggest a role for different modalities in this process, existing studies examining attention effects in syntactic choice are primarily based on visual cueing paradigms. Hence, it remains unclear whether attentional effects on syntactic choice are limited to the visual modality or may be subject to cross-modal interaction. The current study addressed this issue. Native English participants viewed and described line drawings of simple transitive events while their attention was directed to the location of the agent or the patient of the depicted event by means of either an auditory (monaural beep) or a motor (unilateral key press) lateral cue. Our results show an effect of cue location, with participants producing more passive-voice descriptions in the patient-cued conditions. Crucially, this cue location effect emerged in the motor-cue but not in the auditory-cue condition, as confirmed by a reliable interaction between cue location (agent vs. patient) and cue type (auditory vs. motor). Our data suggest that attentional effects on the speaker's syntactic choices are modality dependent and appear to be more prominent in the visuomotor domain than in the auditory domain.

Keywords: visual, auditory, motor, attention, sentence production, syntactic choice, priming.

JEL Classification: Z

¹ Centre for Cognition and Decision Making, National Research University Higher School of Economics, Russian Federation, mpokhoday@hse.ru

²Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK, christoph.scheepers@glasgow.ac.uk

³ Centre of Functionally Integrative Neuroscience (CFIN), Department of Clinical Medicine, Aarhus University, Denmark; National Research University Higher School of Economics, Russian Federation, yury.shtyrov@cfin.au.dk

⁴ Department of Psychology, Northumbria University, Newcastle-upon-Tyne; Centre for Cognition and Decision Making, National Research University Higher School of Economics, Russian Federation, andriy.myachykov@northumbria.ac.uk

⁵ Funding: The study has been funded by the Russian Academic Excellence Project '5-100'.

Introduction.

Multimodal Attention and Syntactic Choice

Describing visually perceived events requires a number of decisions by the speaker, including the selection of words, assignment of thematic roles, and specification of syntactic structure. One aspect of those decisions is structural, such as the specific syntactic choice between active and passive voice in English sentence production. Multiple studies (see Tomlin & Myachykov, 2015, for a recent review) showed that directing the speaker's attention to one of the event's referents by means of *visual* cueing biases subject role assignment to the cued referent, leading to the subsequent choice between the available structural alternatives. For example, speakers tend to select the active-voice frame when their attention is cued towards the event's agent and they tend to select the passive-voice frame when their attention is cued to the patient. It remains unclear, however, whether this correspondence between attentional focus and syntactic choice is limited to the visual modality and, correspondingly, to the use of visual cues. At least hypothetically, the underlying link between attention and grammar may be more general and, therefore, modality-independent. The latter is intuitively plausible for two reasons. First, in real-life conversations, we tend to not only talk about visually perceived aspects of described events. In fact, relevant sensory information about an event can originate from any sensory modality or from multiple modalities simultaneously. The resulting syntactic choice may therefore reflect an integrated attentional bias resulting from signals from different sensory inputs. Second, studies on multimodal integration show convincingly that inputs originating from different sensory modalities converge in an interactive manner (e.g., Giard & Peronnet, 2006; Macaluso, Frith, & Driver, 2000; Molholm, et al., 2002). Therefore, it is possible that attentional cues originating from other modalities can lead to a similar mapping between the cued referent and its constituent role in the sentence's syntactic structure. Here, we present results of a study, in which native speakers of English engaged in a picture-description sentence production task while their attention was directed towards one of the two protagonists of a transitive event by means of *motor* and *auditory* lateral cues. Before introducing the study in more detail, we will briefly review the available evidence.

Visual attention and sentence production

Several existing studies used variants of the visual cueing paradigm (Posner, 1980) in order to demonstrate that syntactic choice, at least in part, reflect the distribution of the speaker's visual attention across the elements of the described scene (Anton-Mendex, 2017; Gleitman, et al., 2007; Hwang & Kaiser, 2009, 2015; Ibbotson, et al., 2013; Montag & MacDonald, 2013; Myachykov, et al., 2011, 2012; Tomlin, 1995, 1997). A recurrent finding in these studies is that the visually cued element tends to be selected as the syntactically most prominent sentence constituent (e.g., Subject),

and thereby trigger the corresponding syntactic choice. In one study demonstrating this effect (Tomlin, 1995), participants viewed and described an animation film depicting one fish eating another fish. Participants' attention was directed to one of the fish by means of an explicit (i.e., consciously processed) visual pointer presented above it. When participants' attention was directed to the agent fish, active voice sentences (e.g. *the blue fish ate the red fish*) were by far the most likely types of descriptions; when attention was directed to the patient fish, passive voice descriptions (e.g. *the red fish was eaten by the blue fish*) were predominantly articulated. The authors concluded that the (explicitly) visually cued referents were more likely to be assigned the subject position in an English transitive sentence.

In another study by Gleitman et al (2007), participants described visually presented events after an implicit visual cue (brief pointer flash, not reaching the speaker's awareness) that directed their attention to one of the subsequently presented referents' locations. Once again, participants were more likely to produce sentences in which the cued referent appeared in the sentence-initial slot, acting typically as its Subject. Importantly, the findings from these and similar studies are corroborated by evidence from neuroimaging studies also showing a direct link between the allocation of attention to the event's referents and the corresponding constituents' grammatical roles and features (Caffarra, Pesciarelli, & Cacciari, 2013; Iwabuchi et al., 2013). Finally in a study by Myachykov et al. (2012), speakers' syntactic choice was analyzed under combinations of syntactic, lexical, and visual priming manipulations. Adding to the previous studies, visual cueing effects were registered both in the presence and in the absence of concurrent linguistic manipulations (syntactic priming and lexical priming) suggesting that syntactic choice may be an integral product of information biases derived simultaneously from different information inputs. This accumulating psycholinguistic evidence is further corroborated by neural simulations and neuroimaging studies showing an intimate link between attention and naming (e.g., Garagnani, Wennekers, & Pulvermüller, 2008; Rommers, Meyer, & Praamstra, 2017)

Multimodal and cross-modal attention

The evidence reviewed above provides important support to the idea that referent-directed visual attention plays a distinct role in the syntactic selection process. At the same time, psycholinguistic research on the role of attention in sentence production in general, and in motivating syntactic choice in particular, has been limited to manipulations of visual spatial attention and, correspondingly, to the use of the visual cueing paradigm. While the visual modality undoubtedly dominates during event perception, adequate understanding of a perceived event has to rely upon input from other perceptual modalities as well (e.g., auditory and motor). As a result, attentional contribution to the sentence production process may result from a selective process that

filters information from other perceptual modalities in addition to the visual one. From this point of view, the attentional system can collect available sensory input from individual modalities and integrate it as a general attentional bias toward one of the available syntactic frames.

This “generalist” view of attention is supported by a number of studies on cross-modal attention showing that spatial attention can be successfully manipulated by e.g. tactile (Spence and McGlone, 2001; Ho et al., 2005, 2006; Scott and Gray, 2008; Yates and Nicholls, 2009, Spence, 2010) and auditory cues (Spence and Driver, 1994; Ho and Spence, 2005; Santangelo and Spence, 2007; Kostov and Janyan, 2012). For example, a recent study by Kostov and Janyan (2012) showed that a lateral *auditory* cue modulated the emergence of a lateral affordance effect that has been previously repeatedly shown to emerge from *visual* perception of manipulable objects. Instead of using visual cues, Kostov and Janyan (2012) cued their participants towards a manipulable object’s (e.g. kitchen appliances) handle with an auditory cue. The cue was either congruent or incongruent with the direction of the object’s handle. Participants gave faster responses when the location of the cue and the direction of the handle were congruent with the response button. Similarly, a study by Santangelo and Spence (2007) showed that unimodal (visual and auditory), as well as multisensory (audiovisual) attentional cues are both capable in capturing visuo-spatial attention. Furthermore, similar results have been obtained for a combination of auditory, tactile and audio-tactile cues (Spence and Santangelo, 2009). These and other studies clearly show that auditory and motor cues as well as visual cues can successfully orient spatial attention.

On the other hand, neuroimaging studies also clearly show that the human brain simultaneously processes inputs from multiple input modalities. Attention in this case plays the role of a filter, deciding which sensory input should be prioritised (Carassco, 2011). This system integrates both the top-down control processes, such as the current goals, but also bottom-up processes, such as the saliency of the perceived stimuli (e.g., Koch and Ulfman, 1999, for a description of a saliency map and its role in shifts of visual attention). Importantly, the strength of the resulting attentional bias may differ depending on the specific modality of the sensory input. Animal studies confirm this *asymmetry* of the interplay between visual attention, on the one hand, and auditory and motor attention systems, on the other. More specifically, they suggest that there is a more intimate link between visual and motor attention than between visual and auditory attention (Goldberg & Segraves, 1987). TMS and fMRI studies (see Iacoboni 2006, for a review) present evidence that the organisation of the posterior parietal cortex of the human brain is quite similar to that of primates, thus suggesting that the relation between attentional modalities in humans is very similar.

Indeed, a number of neuroimaging studies confirm the existence of a highly automatic

multifaceted attentional system based on integrating input from different modalities (e.g., Koelewijn, Bronkhorst, & Theeuwes, 2010; Talsma, et al., 2010; Shinn-Cunningham, 2008). This is further supported by ample neuroimaging evidence confirming the existence of mechanisms for multimodal sensory integration in the human brain (e.g., Kong, et al., 2012; Macaluso, Frith, & Driver, 2000). Put together, the psycholinguistic and the neuroimaging evidence reviewed thus far points to the possibility that there can be a single cross-modal attentional mechanism underlying syntactic choice during sentence production. If so, then both motor and auditory attentional cues should influence syntactic choices in a similar fashion as has previously been documented for visual attentional cues. This suggestion, however, has not been experimentally tested so far.

To address this gap and test the hypothesis of multi-modal attention in syntactic choice, the present study examined how auditory and motor lateral cues affect syntactic choice during transitive sentence production in English. English speaking participants described extemporaneously visually presented transitive events involving two referents, an agent and a patient. Their attention was directed towards the location of one of the interacting referents by means of either an auditory or a motor cue. We analysed the proportion of passive-voice responses as a function of cue location and cue modality. Of particular interest was whether attentional cues from the motor and auditory domain influence syntactic choices in a manner comparable to visual attentional cues in previous research) or whether there are any asymmetries in the effectiveness of different types of cues on syntactic choices.

Method

Design

Two factors were orthogonally manipulated: Cue Type (the cue was provided in the *auditory* or in the *motor* modality) and Cue Location (attention was directed towards the *agent* or towards the *patient* of the depicted event for description). This resulted in a 2x2 factorial design with Cue Type and Cue Location as within-subjects/within-items factors. The dependent variable was the proportion of passive voice event descriptions.

Participants

24 participants (16 females, mean age = 21.7, SD = 2.87) were recruited from the population of students and staff at Northumbria University. To participate in the study, participants had to be monolingual native English speakers, have normal (or corrected to normal) vision, and have no language or attention-related disorders (e.g., dyslexia, ADHD). Participants received course credits

or £5 subject remuneration for their participation. All participants gave written informed consent before taking part.

Materials

Picture materials including character names, event names, stimuli and filler pictures were adopted from the study by Myachykov et al. (2012). Target pictures depicted six transitive types of events (*hit, shoot, chase, touch, push, kick*) rotated across sixteen referents (*artist, chef, clown, cowboy, monk, nun, pirate, policeman, swimmer, dancer, professor, waitress, burglar, boxer, and soldier*). Eight pairs of different characters interacted in each of the six event types, giving 48 transitive-event pictures in total (for an example see Figure 1). There were equal numbers of agent-on-the-left and agent-on-the-right event orientations. Also included were 98 filler pictures, depicting non-transitive events. Materials were presented in a pseudo-random order such that a minimum of two filler pictures separated the target pictures from each other. Materials were arranged into four lists, which allowed all events to feature in all four experimental conditions.

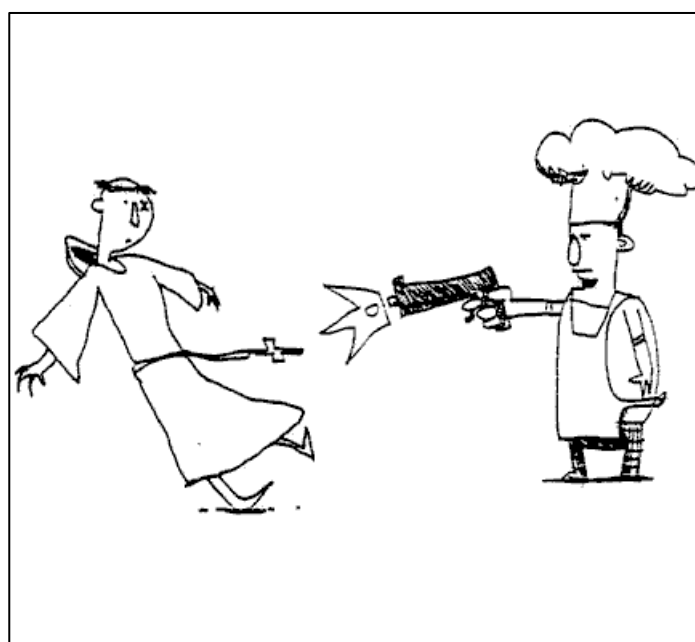


Figure 1. Stimulus example: a transitive event “The monk is shot by the chef”.

Apparatus

The experiment was implemented in SR Research Experiment Builder v1.6.121 (SR Research Ltd, Ottawa, Canada). An EyeLink 1000 Desktop eye-tracker (SR Research) was used to control the participant’s gaze allocation. A Clear view 17-inch display with a refresh rate of 60Hz was used. Eye movements were recorded from the right eye only with a 1000 Hz sampling rate. Auditory cues were delivered by Genius E120 Desktop Speakers via ASIO sound card. Generated sentences were

recorded using a voice recorder application (“Voice Record PRO v.3.2.2” for Android) and stored on a password protected PC. Participants were seated 60 cm away from monitor with their head position controlled by a chinrest. Participants were instructed not to force their head on the chinrest, so they could speak normally.

Procedure

The study took place in the eye-tracking laboratory on Northumbria University campus. On arrival, participants provided their demographics and signed consent forms. After reading experimental instructions, participants first went through a practice session and calibration (5-10 minutes depending on the ease of calibration). During the practice session, participants were engaged in two tasks. First, they familiarized themselves with the 16 referents that were to appear in the main experiment: The character’s depictions were sequentially presented centrally on screen, with their names written underneath. Participants simply read the character names aloud. This ensured that participants familiarized themselves with the referents’ appearances and names, and minimized the chances that participants would find the referents difficult to recognize and describe. Also, this procedure reduced potential ambiguity in naming referents (e.g., *monk*, *priest*, *clergyman* – for the character “monk”). Second, participants practiced describing events similar to the ones they would later encounter in the main experimental session. Participants saw fourteen randomly selected events in a randomized order, with each picture depicting an event with one or two referents (previously practiced) and its name (e.g., *chase*) written underneath. As before, participants were instructed to examine the event and read its name aloud. The purpose of the event practice session was to minimize the variability of potential lexical candidates for the event description (e.g., *punch*, *beat* for the “hit” event), especially since previous research has shown that different verbs have different passive-vs-active voice distributions in corpora (Ferreira, 1994). Also, participants were instructed to keep descriptive detail in their responses to a minimum (i.e. to produce simple descriptions like “The pirate kicks the boxer” rather than “The pirate with one leg forcefully kicks the boxer, and the boxer will probably fall over because of that”).

Upon completion of the practice session, participants received instructions for the main part of the experiment. Participants were instructed that they will be presented with either an auditory cue (AC, auditory signal) or a motor cue (MC, coloured circle), respectively.

In AC trials, an auditory signal (a ‘beep’ of 500 ms duration) was presented laterally either to the left or to the right of the participant. To achieve maximum correlation between visual location of the cued referent and the source of the cue itself (Spence, 2010), the auditory cue originated from speakers that were placed to the left and to the right of the monitor. In AC trials, participants were

instructed to direct their gaze to the fixation circle (black dot in the middle of the screen) and wait. As soon as fixation was confirmed by the fixation trigger (gaze contingent offset after 300 ms fixation maintenance), the lateral auditory cue was played.

In MC trials, participants were instructed to press the green or the blue key depending on the color of a central circle that was presented after the central fixation dot. These two colors were assigned, in a counterbalanced fashion, to two keyboard keys. The ‘Z’ key was used in the leftward cue condition (to be pressed with the index finger of the left hand) and the ‘NumPad2’ key for the rightward cue condition (to be pressed with the index finger of the right hand). Participants were instructed to rest their index fingers on these buttons to avoid visual search for the key locations. For half of the participants, if the colour of the central circle was green (colour code #22B14C), participants had to press the left (‘Z’) key to progress, if it was blue (colour code #3F48CC), they had to press the right (‘NumPad2’) key. For the other half, this mapping rule was reversed. This was followed by the target event presentation (see Figure 2 for a visualization of a trial).

The main task was to describe the event in one sentence mentioning both characters and their interaction. On completion of each trial, participants proceeded to the next trial by pressing Space.

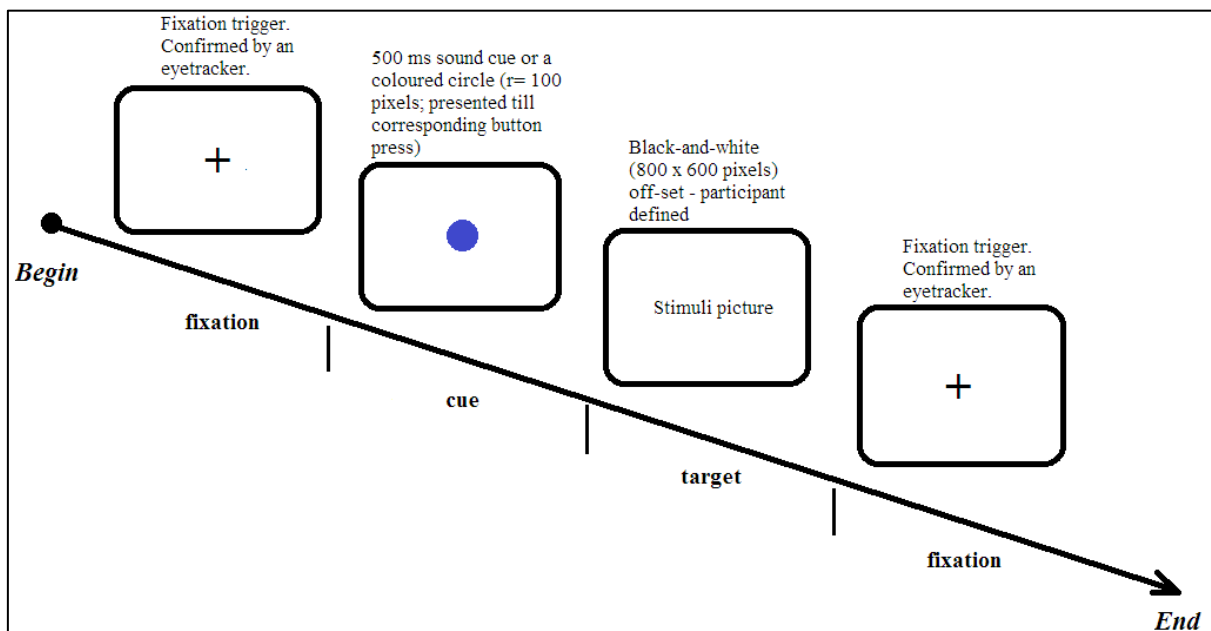


Figure 2. A diagram of the experimental procedure.

Data analysis

The audio-recordings of the event descriptions were transcribed and coded as either Active Voice or Passive Voice. To be coded as Active Voice, the description had to include a subject noun phrase referring to the agent, followed by a transitive verb describing the event, and an object noun phrase

referring to the patient (e.g. *The pirate is chasing the boxer*). To be coded as Passive Voice, the description had to include a subject noun phrase referring to the patient, followed by a passivized transitive verb describing the event, and a by-phrase referring to the agent (e.g. *The boxer is being chased by the pirate*). Conjoined noun phrases (e.g. *The boxer and the pirate are running*), truncated passives (e.g. *The boxer is being chased*), incomplete or ungrammatical sentences, and multi-sentence responses were excluded from the analysis, affecting less than 1% of the data.

Inferential analyses were performed using Generalized Linear Mixed Effects Models (GLMM), as part of the `lme4` package in R. The dependent variable of interest was the occurrence of a Passive Voice description (True = 1, False = 0), and therefore a *binary logistic* model was specified in the family argument of the `glmer()` function. The model included a full-factorial Cue Modality (Auditory, Motor) \times Cue Location (Agent, Patient) fixed effects design. All predictors were mean-centred using deviation-coding. Following Barr, Levy, Scheepers, and Tily (2013), the maximal random effects structure justified by the design was used. Since the two experimental variables were both within-subjects and within-items, the model therefore included by-subject and by-item random intercepts as well as by-subject and by-item random slopes for every main effect and interaction term in the fixed effects design. Random correlations were also included. P-values were obtained via Likelihood Ratio Chi-Square ($LR\chi^2$) model comparisons.

Results

Table 1 shows the distribution of active versus passive voice responses across experimental conditions. As can be seen from the table, particularly in the motor modality, patient-cueing resulted in increased proportions of passive voice responses at the expense of active voice responses.

Table 1. Probabilities of Active versus Passive Voice responses across all participants and trials (absolute cell counts in brackets) by levels of Cue Modality (Auditory, Motor) and Cue Location (Agent, Patient).

Cue Modality	Cue Location	Active Voice	Passive Voice
Auditory	Agent	.750 (432)	.250 (144)
	Patient	.743 (428)	.257 (148)
Motor	Agent	.752 (433)	.248 (143)
	Patient	.649 (374)	.351 (202)

The grand average intercept of the GLMM was estimated as -1.325 log odds units ($SE = 0.251$), which is considerably smaller than zero (i.e., smaller than 0.5 in probability space). This is because passive-voice responses were greatly outnumbered by active-voice responses overall, a result that is in line with previous experimental findings (see Myachykov & Tomlin, 2015, for a

review) as well as with corpus-data (e.g., Roland et al, 2007). In terms of experimental influences, there was a reliable main effect of Cue Location ($LR\chi^2(1) = 5.299, p = .021$) due to more passive-voice descriptions in the Patient-Cued than in the Agent-Cued condition. While for all conditions together the main effect of Modality was not significant ($LR\chi^2(1) = 0.641, p = .423$), there was, crucially, a reliable two-way interaction ($LR\chi^2(1) = 4.940, p = .026$), indicating that the effect of Cue Location was dependent on Cue Modality (see Figure 3). Simple effect analyses (based on treatment-coding of the Cue Modality predictor) confirmed that the effect of Cue Location was significant only in the Motor modality ($b = .743, SE = .256; LR\chi^2(1) = 8.010; p < .005$) but not in the Auditory modality ($b = .021, SE = .193; LR\chi^2(1) = 0.012; p = .914$). Therefore, it can be concluded that attentional cues in the motor modality, but not in the auditory modality affected participants' syntactic choices for event description.

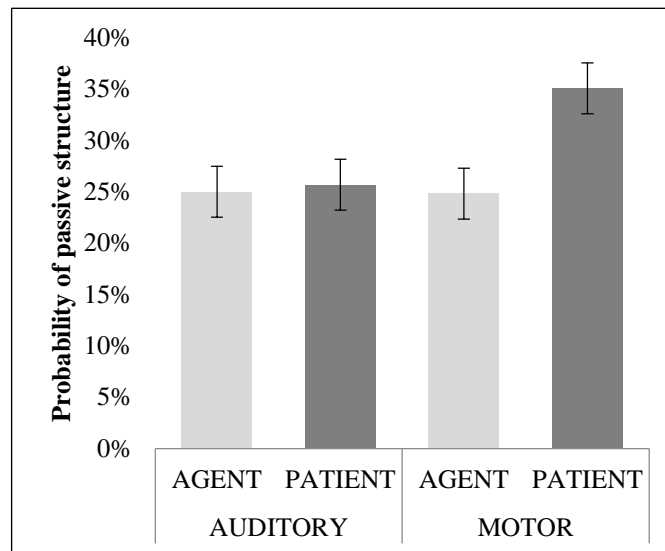


Figure 3. Group-average probability of passive-voice selection as a function of Cue Location

Discussion

The aim of this study was to investigate whether motor and auditory lateral attentional cues affect syntactic choice during visually mediated sentence production. Previous research has shown that manipulation of attention by visual modality cues leads to the assignment of the cued referent to the Subject position in a spoken sentence, thus triggering the choice of syntactic structure. Furthermore, neuroimaging and cognitive studies suggest that attentional modalities have common control mechanisms, but due to differences in processing might have variations in the way they affect syntactic choice. Overall, our results show that *motor* cues affect syntactic choice in a manner similar to visual cues, indicating the existence of a partially multimodal attentional mechanism affecting syntactic choice in sentence production. However, the same was not true for the auditory cues. Below, we discuss the implications of this main finding.

First, the results of the present study provide further support to the existing research on the role of attention in biasing syntactic choice during sentence production. Our data show that participants were more likely to describe target events by using passive-voice sentences when the previously presented cue was in the patient location. This replicates the pattern of results from studies of visual attention and sentence production studies (cf. Gleitman et al., 2007; Myachykov et al., 2012; Tomlin 1995, 1997). Thus, cueing attention before sentence production is an important factor affecting the grammatical structure of the sentence.

Secondly, our data reveal, for the first time, that this effect is not limited to the visual modality, since priming the speaker's attention via lateral motor cues led to a similar priming effect, as confirmed by a simple-effect analysis of the reliable interaction between Cue Type and Cue Location. The same analysis revealed that the main effect of Cue Location was carried almost exclusively by motor-modality cues, while proportions of passive-voice descriptions in the auditory-cue trials were not reliably affected by Cue Location. This suggests constraints on the scope of any cross-modal components in the interface between attention and syntactic choice. One further possibility is to understand whether the same pattern of results can be achieved in less strict word-order languages (e.g. Russian, Finnish, etc.). There have been studies of visual cueing in Russian (Myachykov and Tomlin, 2008), Korean (Hwang and Kaiser, 2009) and Finnish (Myachykov et al, 2011) which show somewhat conflicting results. Whereas in Russian, syntactic choice was biased by visual cueing, no syntactic structure alterations were observed in Korean and Finnish, although the attentional manipulations were successful as such. Possibly, other attentional modalities can have a similar effect on syntactic choice in different languages.

Finally, the difference between the auditory and the motor cues' capacity to affect syntactic choice in our study may result from the intrinsic differences between the auditory and motor cues we used. The capacity to laterally displace attention is generally different for these two modalities with auditory cues being less powerful in this respect (Posner, 1978). This is partially explained by the fact that in the auditory modality, the location of the cue source is extracted from the signal differences between the two cochleas in the tonotopically organized auditory cortex making the successful cue processing easier (Rayleigh, 1907). Conversely, in order to process a motor cues similar to the ones used in our study, one has to process the color of the cue and then make a decision regarding the required button press before executing the action of pressing. As a result, the auditory cue may be less successful in biasing syntactic choice simply because it requires less processing and it does not lead to the establishment of a strong lateral attentional bias. Furthermore, as neuroimaging studies suggest, although some control mechanisms are shared between cues from different attentional modalities, orienting mechanisms differ. This opens a possibility to further

study the relation between attentional modalities. Arguably, it is possible to shut down or enhance orienting of attention; for example, repetitive TMS of right angular gyrus caused interruptions in orienting of attention (Rushworth et al., 2001), while a 1Hz synchronized TMS of left or right parietal cortex for 10 minutes significantly improved subjects' attention to ipsilateral targets (Hilgetag, et al., 2001). Thus, the present hypothesis may be further tested by using TMS in languages where there is seemingly no effect of cueing on syntactic choice, such as Finnish or Korean. Stimulation or inhibition of the above brain regions can possibly cause speakers of such languages to switch syntactic structure towards the cued referent and thus it can help generalizing this mechanism between languages with different levels of strictness of word order.

Conclusions

To describe a picture in a particular way, an interplay between a number of linguistic and nonlinguistic factors needs to be established. One of those factors is how attention is allocated towards the referents of the described event. Our study has shown for the first time that not only by manipulating visual attention can we affect this process. Motor attention (e.g. lateral button presses) can bias the choice of syntactic structure by shifting attention towards the referent. In contrast, an auditory cue (a lateralized sound) did not generate the same effect, which may be due to a number of factors, such the depth or time of cue processing or a closer link between visual and motor networks in the brain. Future research is necessary to address the multi-modal attention mechanisms in syntactic choice in more detail, as well as to test their generalizability to other languages and to establish their underlying neural mechanisms.

References

- Anderson, S. J., Yamagishi, N., & Karavia, V. (2002). Attentional processes link perception and action. *Proceedings of the Royal Society of London B: Biological Sciences*, 269(1497), 1225-1232.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3), 255-278.
- Caffarra, S., Pesciarelli, F., & Cacciari, C. (2013). The interaction between language and visual spatial attention systems in grammatical gender processing. An N2pc study. *Cognitive neuroscience*, 4(3-4), 217-224.
- Hilgetag, C. C., Théoret, H., & Pascual-Leone, A. (2001). Enhanced visual spatial attention ipsilateral to rTMS-induced 'virtual lesions' of human parietal cortex. *Nature neuroscience*, 4(9), 953-957.
- Ibbotson, P., Lieven, E. V., & Tomasello, M. (2013). The attention-grammar interface: eye-gaze cues structural choice in children and adults. *Cognitive Linguistics*, 24(3), 457-481.
- Iwabuchi, T., Inui, T., Ohba, M., & Ogawa, K. (2013). A functional MRI study of a picture–sentence verification task: evidence of attention shift to the grammatical subject. *NeuroReport*, 24(6), 298-302.
- Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language*, 33(6), 715.
- Gleitman, L.,R., January,D., Nappa,R. and Trueswell, J.C. (2007). On give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, 57, 544-569.
- Goldberg, M. E., & Segraves, M. A. (1987). Visuospatial and motor attention in the monkey. *Neuropsychologia*, 25(1), 107-118.
- Ho, C., Tan, H.Z., Spence, C. (2005). Using spatial vibrotactile cues to direct visual attention in driving scenes. *Transportation Research Part F* 8, 397–412.
- Ho, C., Reed, N.J., Spence, C. (2006). Assessing the effectiveness of “intuitive” vibrotactile warning signals in preventing front-to-rear-end collisions in a driving simulator. *Accident Analysis & Prevention* 38, 989–997.
- Ho, C., & Spence, C. (2006). Verbal interface design: Do verbal directional cues automatically orient visual spatial attention?. *Computers in Human Behavior*, 22(4), 733-748.
- Hwang, H., & Kaiser, E. (2009). The effects of lexical vs. perceptual primes on sentence production in Korean: an on-line investigation of event apprehension and sentence formulation. In *Talk presented at the 22nd CUNY conference on sentence processing, Davis, CA*.
- Kong, L., Michalka, S. W., Rosen, M. L., Sheremata, S. L., Swisher, J. D., Shinn-Cunningham, B.

- G., & Somers, D. C. (2012). Auditory spatial attention representations in the human cerebral cortex. *Cerebral Cortex*, *24*(3), 773-784.
- Kostov, K. & Janyan, A. (2012). The role of attention in the affordance effect: can we afford to ignore it? *Cognitive Processes*, *13*, 215-218.
- Montag, J. L., & MacDonald, M. C. Choice of Referring Expression Subject to Visual Context, Linguistic Context and Competition.
- Myachykov, A., Garrod, S., & Scheepers, C. (2012). Determinants of syntactic choice in visually situated sentence production. *Acta psychologica*, *141*(3), 304-315.
- Myachykov, A., Thompson, D., Scheepers, C., & Garrod, S. (2011). Visual attention and syntactic choice in sentence production across languages. *Language and Linguistics Compass*, *5*(2), 95-107.
- Posner, M. I. (1978). Chronometric explorations of mind. Lawrence Erlbaum.
- Posner, M.I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3-25.
- Rayleigh, L. (1907). XII. On our perception of sound direction. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *13*(74), 214-232.
- Rushworth, M. F., Ellison, A., & Walsh, V. (2001). Complementary localization and lateralization of orienting and motor attention. *Nature neuroscience*, *4*(6), 656-661.
- Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of memory and language*, *57*(3), 348-379.
- Santangelo, V., & Spence, C. (2007). Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(6), 1311.
- Scott, J.J., Gray, R. (2008). A comparison of tactile, visual, and auditory warnings for rear-end collision prevention in simulated driving. *Human Factors* *50*, 264-275.
- Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in cognitive sciences*, *12*(5), 182-186.
- Spence, C. (2010). Crossmodal spatial attention. *Annals of the New York Academy of Sciences*, *1191*(1), 182-200.
- Spence, C., McGlone, F.P. (2001). Reflexive spatial orienting of tactile attention. *Experimental Brain Research* *141*, 324-330.
- Spence, C., & Santangelo, V. (2009). Capturing spatial attention with multisensory cues: A review. *Hearing research*, *258*(1), 134-142.
- Tomlin, R. S. (1995). Focal attention, voice, and word order: an experimental, cross-linguistic

- study. *Word order in discourse*, 517-554.
- Tomlin, R. S. (1997). Mapping conceptual representations into linguistic representations: The role of attention in grammar. *Language and conceptualization*, 162-189.
- Tomlin, R.S. & Myachykov, A. (2015). Attention and salience. In E. Dabrowska & D. Divjak (Eds.) *Handbook of Cognitive Linguistics*. (pp. 31-53). De Gruyter Mouton.
- Yates, M.J., Nicholls, M.E.R., 2009. Somatosensory prior entry. *Perception & Psychophysics* 71, 847–859.
- Macaluso, E., Frith, C. D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science*, 289(5482), 1206-1208.
- Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive Brain Research*, 14(1), 115-128.
- Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of cognitive neuroscience*, 11(5), 473-490.
- Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends In Cognitive Sciences*, 14(9), 400-410.
- Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica*, 134(3), 372-384.

Mikhail Y. Pokhoday

National Research University Higher School of Economics (Moscow, Russia). Centre for Cognition and Decision Making, research assistant.

E-mail: mpokhoday@hse.ru

Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.

© Pokhoday, Scheepers, Shtyrov, Myachykov, 2017