A perceptual learning approach to the Whorfian hypothesis: Supervised classification of motion.

Panos Athanasopoulos[1] and Daniel Albright[2]

[1]Department of Linguistics and English Language, Lancaster University

[2]School of Psychology and Clinical Language Sciences, University of Reading

ABSTRACT

Recent research on the relationship between grammatical aspect and motion event construal shows that speakers of non-aspect languages (e.g. German, Swedish) attend to event endpoints more than speakers of aspect languages (e.g. English, Spanish) in non-verbal categorization tasks (Athanasopoulos & Bylund, 2013; Flecken et al., 2013; von Stutterheim et al., 2012). In this paper we take a perceptual learning approach to the Whorfian hypothesis, training native speakers of English to categorize events either in an English-like way (same-language bias) or in a Swedish-like way (other-language bias), with and without verbal interference in English. Results showed that successful learning occurred in both language conditions, and that verbal interference disrupted learning only in the condition where the perceptual dimension to be learned was also salient in the participant's native language, but not in the condition where it was not. This suggests that individuals may recruit verbal processes online for the purposes of classification more readily when the stimuli to be classified are also habitually encoded in the native language, but may rely on language-independent perceptual mechanisms when learning to classify stimuli not habitually encoded in their native language.

Keywords: Motion Events, Supervised Classification, Whorf, Linguistic Relativity

1. Introduction

Since the 1940s, when it first came to prominence, the idea of linguistic relativity has been a controversial one. Although the idea has been restated and reformulated over the years, Whorf's original formulation of the idea was a simple one: ''users of markedly different grammars are pointed by their grammars toward different types of observations and different evaluations of externally similar acts of observation'' (Whorf, 1956 [1940], p. 221). Early relativity researchers operationalized Whorf's formulation into an empirically testable research program, looking at the degree of verbal mediation of categorical perception processes. By definition, categorical perception entails that the categories possessed by an observer influence the observers' evaluation of perceptual phenomena that are otherwise stable and unchanging, for instance perceiving a perceptual continuum like colour as divided into discrete categories, which we call 'blue', 'green', etc. (see Goldstone & Hendrickson, 2009, for an advanced treatment). The debate then focused on whether those categorical boundaries are innate and universal (the universalist position, see Berlin & Kay, 1969; Rosch-Heider, 1972; and Franklin, Clifford, Williamson & Davies, 2005, for a modern version), or whether those categorical boundaries come from the native language of the observer and vary as a function of linguistic diversity (the relativist position, see Brown & Lenneberg, 1954; Kay & Kempton, 1984; and Agrillo & Roberson, 2009, for a modern version).

In recent investigations, however, the way in which linguistic relativity is discussed is often quite different than the way in which Whorf first wrote about it. Some scholars have misinterpreted, or misrepresented linguistic relativity with some form of radical linguistic determinism: if an idea is not linguistically coded in our language, then we cannot ever conceive of that idea (see e.g. Steven Pinker's critical treatment in his 1994 book). Casasanto (2008) among others helpfully demonstrates how such interpretation is far removed from the linguistic relativity hypothesis. The hypothesis of whether we are able to think or not without language is orthogonal to the hypothesis that having a categorical distinction in language makes that distinction more salient in cognition (the Whorfian view). Although the majority of scientific opinion on the matter has left behind this strawman argumentation, such flawed

argumentation still makes its way into popular scientific opinion (see e.g. McWhorter, 2014).

On the other extreme, some scholars equate linguistic diversity (the fact that different languages encode different lexical and grammatical categories) with linguistic relativity (the hypothesis that those categories then affect behavior) and present evidence drawn from the former observation as evidence for the latter hypothesis. The circularity of such reasoning is made obvious by Casasanto (2008), among many others: "the project of inferring cognitive differences solely from linguistic differences is hopelessly circular. Patterns in language can serve as a source of hypotheses about cognitive differences between members of different language communities, but some sort of extralinguistic data are needed to test these hypotheses: Otherwise, the only evidence that people who talk differently also think differently is that they talk differently!" (p. 67). The vast majority of empirical research shows that verbal and non-verbal behaviour are almost never perfectly isomorphic. Examining non-verbal behavior is crucial in establishing the degree of verbal mediation in human thinking processes (Casasanto, 2008; Bylund & Athanasopoulos, 2014). Some scholars (e.g. Pavlenko, 2014) even maintain that Whorf was not interested in non-verbal behavior at all, something which is at odds with Whorf's own formulation of the hypothesis (see above) and the very fact that he developed the hypothesis by observing non-verbal behavior: industrial accidents occurring as a result of people smoking and dropping lit cigarrete butts near gasoline drums labeled 'empty'. An empty gasoline drum contains vapours that are highly flammable. Yet the verbal label 'empty' prompted workers to downplay that potential danger in their non-verbal behavior.

In fact, there has been an increasing number of studies that show that linguistic relativity effects, instead of altering a speaker's "worldview," have a tendency to affect—but not deterministically change—universal perceptual processes. For example, a study by Thierry et al. (2009) shows that both Greek and English speakers are perceptually aware of differences between light and dark blue, and between light and dark and green, indexed by differential brain activation to different degrees of luminance regardless of color. At the same time there is further increased brain activation in Greek speakers for blue rather than green luminance contrasts because Greek has two distinct basic terms to refer to the blue area of color space.

Similarly, to give an example from a domain other than colour, studies show that English speakers preferentially categorize entities on the basis of common shape rather than common material properties, because shape is the best perceptual indicator of countability, a linguistic feature that is obligatorily observed in the English nominal system in terms of grammatical number marking on count, but not mass, nouns. Speakers of languages like Japanese or Yucatec on the other hand, where grammatical number marking is optional or non-existent, show a preference for material as a basis of categorization (Lucy, 1992; Imai & Gentner, 1997). However, when looking at the general pattern of performance, all speakers are in fact aware of the perceptual distinction between objects and substances, because speakers from all language groups tend to match by common shape more often in a matching condition that involves solid object targets, than in a matching condition that involves malleable substance targets (Lucy & Gaskins, 2003; Imai & Mazuka, 2003). Such phenomena are observed not only with static stimuli such as colors and objects, but also with dynamic ones like motion events, a research domain which is the focus of the current investigation.

2. Relativized Motion

A motion event contains at least three pieces of information: an actor, a ground (the background against which the actor is placed) and the motion action. The motion action consists of a number of pieces of information that are conflated into a single concept, as detailed by Talmy (1991). Many types of information can be included in this conflation: path and manner are the most commonly discussed in linguistic relativity literature, but cause, enablement, precursion, and constitutiveness are also mentioned in Talmy's (1991) discussion.

Most of the investigations into linguistic relativity and motion perception have focused on how the conflation patterns of a language encourage speakers to pay attention to the path or manner of an event (in Talmy's verb-framed and satellite-framed languages, respectively) and whether or not this attention leads to effects on non-verbal behaviour; for example, Gennari et al. (2002) looked at whether or not the linguistic encoding of an event by English and Spanish speakers (whose languages direct attention specifically to manner and path, respectively) affected subsequent

similarity judgment and recognition memory tasks; they found an effect of language in some non-verbal tasks, but not others. Papafragou, Massey, and Gleitman (2002), in a study that included English- and Greek-speaking participants, found no difference between speaker groups on non-verbal tasks, an effect interpreted by the authors to show that "the lexical patterning of the specific language did not bleed into subjects' performance in tasks that do not call on the linguistic categories specifically" (p. 213). While these studies provided some conflicting evidence, they pointed to a dissociation between verbal processing and non-verbal cognitive tasks. Later investigations, however, provided evidence that the conflation pattern of a language may have significant effects on non-verbal behaviour (cf. Kersten et al., 2010). A study by Papafragou, Hulbert, and Trueswell (2008) that utilized eye-tracking found no differences in attention allocation or categorization between Greek and English speaking observers upon initial exposure to a scene. However, eye-tracking revealed that participants subsequently allocated attention to aspects of the scene not encoded in their respective languages (Papafragou, Hulbert, & Trueswell, 2008). This 'reverse Whorfian effect' can be interpreted as a linguistic relativity effect, because attention was guided by perceptual elements that were 'new' to the observer's mental construal of the scene. In other words, those elements captured attention because they had not already been implicitly verbalized by participants (path of motion for English speakers, manner of motion for Greek speakers).

The speculation that participants verbally encoded the scenes implicitly was in fact confirmed in a subsequent study by Trueswell and Papafragou (2010), which showed that between-group differences in attention allocation to motion events were abolished under concurrent verbal interference, where participants need to engage language to perform a verbal interference task (e.g., repeating syllables or strings of digits) parallel to the non-verbal cognitive task (e.g., categorizing motion scenes). Such findings demonstrate that when the verbal system is simultaneously engaged in a different task, the ability to rely on verbal resources for the purposes of categorical judgments is reduced. Such effects have also been demonstrated in other domains such as colour categorical perception (Gilbert, Regier, Kay, & Ivry, 2006; Winawer et al., 2007). Taken together, results from studies employing verbal interference point to an online role of language on categorical judgments, a phenomenon captured by Lupyan's (2012; this issue) label-feedback hypothesis.

3. Grammatical aspect and motion

More recently, a different linguistic typology has become of interest to linguistic relativity researchers studying motion events: that of grammatical aspect. Aspect defines the temporal structure of an event, providing information on how the event unfolds through time. For example, an imperfective/progressive aspect indicates that an event is still ongoing (as in "the man is running"), while a perfective aspect indicates completion (as in "the woman had driven to work"). Speakers of all languages are able to encode descriptions of events using different temporal aspects; the crucial difference, however, is whether or not the means of differentiating between those aspects is grammatical or lexical.

Languages that have grammatical means of expressing aspect, like English and Spanish, use prefixes, suffixes, or syntactical means to encode the aspectual status of an event. In English, the *–ing* verbal ending indicates ongoingness, while the Spanish prefixes *–ando/–iendo* perform the same function. In contrast, languages with lexical aspect, like Swedish and German, use lexical items to encode the same information (see e.g. Athanasopoulos & Bylund, 2013). It is important to note, however, that different languages follow different patterns when it comes to grammaticalisation of aspect. Von Stutterheim et al. (2012) provide an illustrative discussion of the difference between aspectual systems, providing the example that Arabic uses a fully grammaticized system for distinguishing between perfective and imperfective aspects, and Russian uses a combination of grammatical and lexical means, while none of the other languages in the study classified as having grammatical aspect had grammatical means for distinguishing between the two in all situations (p. 838–9).

Previous research has shown that this distinction in the grammaticalisation of aspect gives rise to a notable difference in verbal encoding strategies. Languages that do not contain grammatical aspect tend to encode events holistically, in both linguistic and cognitive tasks. In this encoding, the entire event, from inception to completion, is the focus of the encoding. In contrast, languages with grammatical aspect tend to create phasally decomposed encodings of events, in which the focus is placed on a

specific phase of the event. A good visual representation of this distinction, adapted below, is presented in Langacker's (2008) *Cognitive Grammar: A Basic Introduction.*
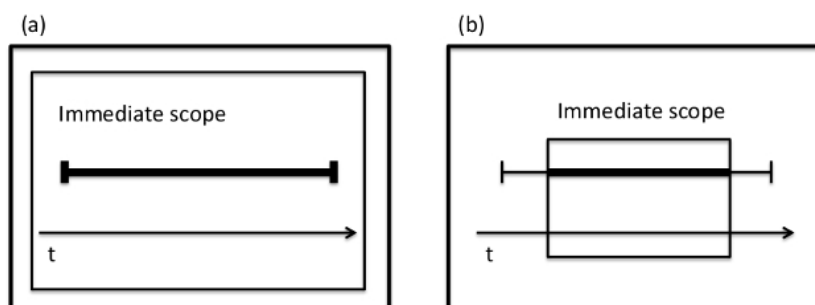

Figure 1. An event is shown by a horizontal line, with inception and conclusion; the part of the event that is emphasised by a particular language is represented by a thicker line. (a) shows the immediate temporal scope of a language with no grammatical aspect (holistic focus including endpoints), while (b) shows the immediate temporal scope of a language with grammatical aspect (focus on ongoingness excluding endpoints).

This phenomenon is well-attested in verbal tasks. For example, Flecken's (2011) study showed that speakers of English, a language with the grammatical means to indicate the progressive aspect, were more likely to encode events with a focus on the ongoingness than German speakers, who portrayed the event holistically. The way in which this difference in linguistic encoding is manifested is often as a focus on motion endpoints, as seen in Athanasopoulos and Bylund's (2013) study on speakers of English and Swedish. In this study, participants viewed twelve video clips of intermediate goal orientation (i.e., in which an actor was headed toward an endpoint, but did not reach it before the end of the clip; the endpoint was, therefore, inferable, but not definitely known). Swedish speakers mentioned the endpoint of the motion in 62% of the video clips, while English speakers mentioned it in only 43% (these tendencies were statistically significantly different from chance).

Such cross-linguistic differences in verbal encoding are corroborated by other sources of evidence using eye-tracking. For example, von Stutterheim et al. (2012) demonstrate that when confronted with the task of describing a motion event scene, speakers of different languages exhibit differences in their allocation of visual attention: speakers of Arabic, English and Spanish tend to fixate the possible goal of motion to a significantly lesser extent than speakers of Dutch and German. Such

diverging patterns of attention allocation in speakers of different languages are captured by what von Stutterheim and colleagues call the Seeing-for-Speaking Hypothesis, according to which a speaker of a language that codes a certain meaning grammatically (in this case the notion of ongoingness) will attend to the relevant feature of a given visual scene, whereas a speaker of a language that only codes this meaning optionally through lexical and phrasal means will be more prone to attend to the endpoints of events (von Stutterheim et al. 2012).

While the linguistic effects of grammatical phenomena related to motion events are well-attested, their potential for affecting non-verbal behaviour is less well-known. The methodological framework of modern research into linguistic relativity is anchored on two basic tenets: First, on the proposal that similarity is the basis of categorisation (Nosofsky, 1986), and second, that categorisation is an essential element of human cognition (Harnad, 1987). Against this backdrop, non-verbal behaviour is often operationalised along a continuum of cognitive tasks with an inherent categorisation component (Lucy, 1997). A classic methodological example of this is the triads-matching task, in which the participant is to match a target stimulus with one out of two stimuli alternates, based on their degree of similarity. Athanasopoulos and Bylund (2013) used a triads-matching task to analyze the potential non-linguistic effects of aspectual systems. The third video in each triad contained a video of a motion event with intermediate goal orientation, as discussed above. The first two videos were presented in random order and varied between high goal orientation, in which an actor reaches a destination (e.g., walks through a door), and low goal orientation, in which there was no inferable endpoint (e.g. a person walking). Participants were asked to categorize video clips under four conditions: one in which the clips were shown sequentially (the 'memory' condition), and one in which they were shown concurrently (the 'online' condition); and both with and without verbal interference. Language was shown to significantly affect the categorization choices made by speakers in the memory condition without linguistic interference. That is, Swedish speakers were more likely to choose the high goal orientation video as more similar to the intermediate goal orientation video than English speakers, who in turn showed an increased preference for the low goal orientation video relative to Swedish speakers. Interestingly, in the online condition and in the memory condition under verbal interference, there was no significant

difference between speaker groups. Similarly to the studies by Papafragou and collegaues discussed earlier, this provides evidence that in the domain of motion events, non-verbal functions are only affected when language may be used as a tool—for example, to help the participant store the videos in short-term memory, or when participants mentally encode the scenes using verbal descriptions.

Furthermore, even in the condition that yielded cross-linguistic differences, the global pattern of results revealed that the dominant preference regardless of the participant's language background was for the low goal orientation alternate. This is because all clips showed ongoing locomotion, and in both the low goal orientation alternate and the intermediate goal orientation target the endpoint was not reached. Thus, like the studies by Thierry et al. (2009), Lucy and Gaskins (2003) and Imai and Mazuka (2003), Athanasopoulos and Bylund (2013) showed that the native language of the observer systematically biases, rather than shapes, this language-independent preference. The authors interpreted this finding as "one that points to an attenuating effect of language on perceptual processes that are likely to be universal and unchanging," and one that "contributes to the emerging view that Whorfian effects are not an all-or-nothing phenomenon" (p. 19).

In another non-verbal study, von Stutterheim et al. (2012) found that speakers of languages with lexical aspect were more likely to remember the endpoints of a motion presented in a video clip when shown a still image of that clip with the endpoint removed—however, they are quick to point out that this memory task took place after verbal encoding, leading to the question of whether or not the act of encoding played a role in affecting the storage of the clips in short-term memory of the participants. Further complicating the issue, there can be disagreement over how to interpret results that seem to point toward the same findings. For example, Athanasopoulos and Bylund maintain that the results of their (2013) study support the findings of Papafragou and Selimis (2010), but hold a different interpretation. Whereas Papafragou and Selimis (2010) believe that their results support a model in which language and motion-event encoding are separable, rendering Whorfian effects transient and not affecting 'deep', 'underlying' representation of motion, Athanasopoulos and Bylund (2013) suggest that habitual processes are at work, and that language can affect language-independent perceptual processes. In this line of reasoning, there is no distinction between 'deep' and 'surface' representations.

Representation is constructed as a function of the relative weight of perceptual-experiential and linguistic information (for further discussions see Bylund & Athanasopoulos, 2014; Casasanto & Lupyan, 2015). This mutual synergy between different kinds of linguistic and non-linguistic information is a phenomenon that is not at all trivial given the ubiquity of the effect across a number of different perceptual domains.

4. The present study: Supervised classification of motion

A study by Kersten et al. (2010) took advantage of a different methodology: supervised classification. In this experiment, participants were asked to categorize animations of bug-like creatures that moved across the screen into one of four categories; after they chose a category, they were given feedback as to whether they had answered correctly. The participants were not informed of the categorization criteria. The experiment used two conditions: one in which the correct dimension of categorization was the manner of the motion, and another in which it was the path of the motion. Interestingly, English speakers performed significantly better on manner-based categorizations, a result that was attributed to the focus on manner in English verbs. English and Spanish speakers performed similarly well on the path categorization task, suggesting that path is a universally salient aspect of motion events. The use of this methodology allowed Kersten et al. to suggest that English speakers, when confronted with a classification task that did not rely on path, would more quickly resort to manner information because their native language habitually draws attention to this dimension of the motion event.

This methodology, however, has never been used to investigate aspectual systems. Currently it is not known to what extent linguistic encoding is a mechanism utilized more readily when the observer's native language also encodes the perceptual dimension in question, or whether all perceptual dimensions are by default linguistically encoded. Disrupting the engagement of language by means of verbal interference is an invaluable tool in assessing the role of verbal mediation in classification processes, and thus the Whorfian hypothesis (Perry & Lupyan, 2013). The current study addresses this gap in the motion event research literature, and will compare supervised classification tasks in which participants are under no

interference or verbal interference in an attempt to determine to what degree implicit verbal processing is used in categorizing events on an aspectual dimension. Two experiments will be detailed: a supervised classification task involving stimuli containing intermediate, high and low degrees of endpoint orientation (see earlier discussion), and a supervised classification task with verbal interference. The classification task will be administered in two conditions: one in which the correct dimension of categorization is the low endpoint alternate (the 'English-like' pattern), and another in which it is the high endpoint alternate (the 'Swedish-like' pattern).

Based on previous findings (Athanasopoulos & Bylund, 2013; Bylund, Athanasopoulos & Oostendorp, 2013), we expect that participants will be more accurate in the low endpoint alternate condition than in the high endpoint condition, because perceptually the low endpoint alternate is more similar to the intermediate endpoint target since in both a goal has not been reached (see also earlier discussion). Based on previous training categorization studies (e.g. Kersten et al., 2010), we expect performance to improve progressively in the experiment not involving verbal interference. We can plausibly put forward three hypotheses with regards to the role of linguistic encoding under verbal interference. If observers draw on linguistic encoding generally, regardless of whether ongoingness is grammaticized in their native language, then we should expect no gradual improvement in performance on the classification task in either the high endpoint or the low endpoint conditions since the observers will not have verbal resources at their disposal to perform the task. We call this the 'maximal verbal mediation hypothesis'. If observers are more prone to draw on linguistic encoding when the perceptual dimensions in question are made salient in their native language, then we should expect performance in the verbal interference experiment to improve only in the condition that is not made salient in their native language, since verbal interference will only affect the condition where language-specific categories are employed. We call this the 'selective verbal mediation hypothesis'. Finally, there is also a possibility that participants are not relying on linguistic resources at all to learn how to categorize the stimuli. In such a case, there should be no effect of verbal interference on classification, that is, classification along the high or low endpoint condition will steadily improve even under conditions of verbal interference. We call this the 'no verbal mediation hypothesis'.

Experiment 1: Supervised classification with no verbal interference

5. Method

5.1 Participants

  The experiment was undertaken by 40 undergraduate psychology students (35 female, 5 male) from the University of Reading who were all monolingual native speakers of English. Their ages ranged from 18 to 22.

5.2 Materials

  The stimuli used in this experiment were inspired by the video stimuli used by von Stutterheim et al. (2012) in their study of motion event perception, and by the triads matching design in Athanasopoulos and Bylund (2013), which utilized stimuli created by von Stutterheim and colleagues. In this study we created a new set of stimuli along similar lines to those used previously. Each video in this set contained one female actor who was pictured walking toward a destination (or not, in the case of low-goal-orientation videos). 19 videos were created, and 5 videos from the von Stutterheim set that met the criteria were also used.

  To determine permissible triads, all videos were checked and matched for the direction of the actor—some videos showed the actors walking away from the camera, while others travelled from one side of the frame to the other. Each triad contained only one type of video, and no triad contained two videos that featured the same actor. This ensured that the visual similarity of each video was not affected by the actor. In all, 23 triads were created utilizing the 24 videos in possible combinations.

5.3 Procedure

  The participants performed a standard triads matching task, in which three videos are seen in sequence (labeled as A, B, and X) and the participant is asked to make a decision after seeing the third video. In this case, the A and B videos were high- or low-goal-orientation (A and B never contained a video of the same type) and the X video was always an intermediate-goal-orientation video. Once the X video had

finished playing, two buttons appeared on the screen, labeled "A" and "B." Participants were instructed to click on the button labeled with the letter designating the video that was more similar to the X video. Participants were instructed to focus on what was happening in the videos, and not on environmental factors, like clothing or scenery.

Each triad set was presented twice, so that each triad was seen four times: twice in an ABX order, and twice in a BAX order. Thus there were 96 trials overall.

After clicking on a button, the participants received feedback in the form of a green tick mark over the answer that they clicked on if it was correct, or a red X over the correct answer if they chose incorrectly (they were instructed in this feedback, and verbally acknowledged their understanding of it, before the experiment began).

In condition 1, the correct answer was always the low-goal-orientation video; this is termed the English-typical pattern based on the results of Athanasopoulos and Bylund's (2013) study. In condition 2, the opposite was true, and participants needed to select the high-goal-orientation video (the Swedish-typical pattern). Participants were randomly allocated to one of two conditions such that there were 20 participants per condition, and they were not informed of the classification criteria before the experiment.

6. Results

Although the 96 trials were presented without any intermission, for the purposes of analysing each participant's improvement over time, the trials completed were split into four blocks of 24 trials each. The dependent variable in this case is the proportion of correct answers chosen by the participant in each block, with $1.0 = 24$ correct answers.

Mauchly's test of sphericity was significant, $W (5) = .521, p < .001$, so the Greenhouse-Geisser adjusted F is reported in the following results. There was a significant main effect of block, $F (2.06, 111) = 27.03, p < .001$, indicating a significant improvement across blocks. There was also a main effect of condition, $F (1, 37) = 7.10, p < 0.05$, indicating that participants performed better in the English-typical pattern than the Swedish-typical pattern. There was no significant interaction between block and condition, showing that the rate at which participants learned the patterns were similar.
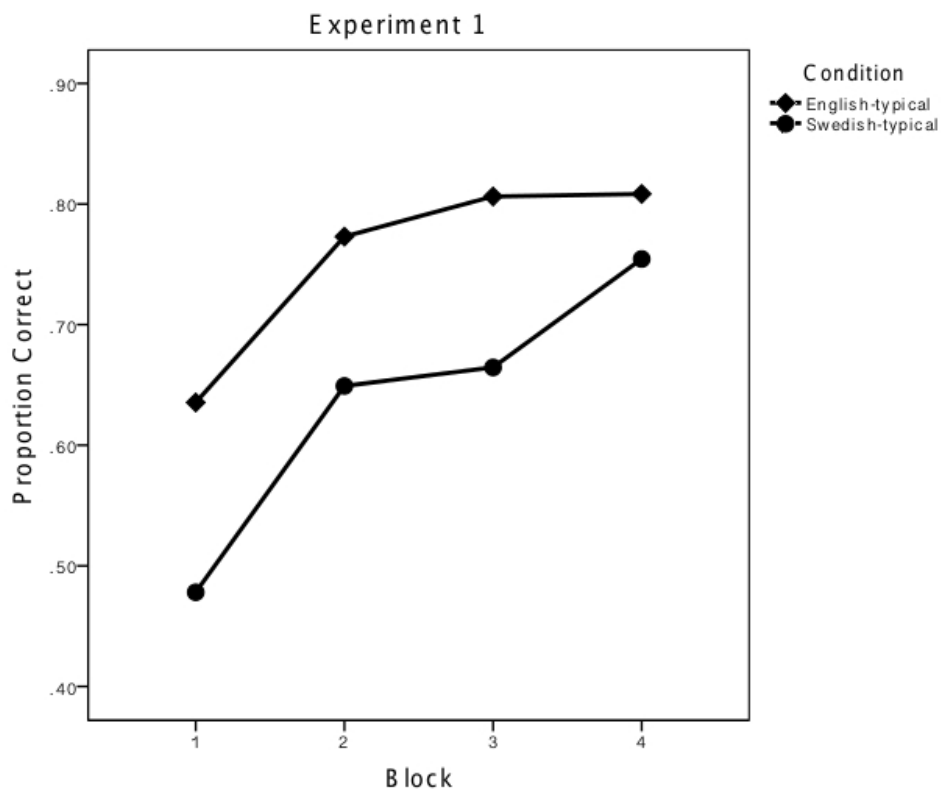
Figure 2. Results of motion video categorisation task in experiment 1. Participants in the English-typical condition scored higher across all blocks than those in the Swedish-typical condition, and both groups showed improvement as they progressed through the trials.

A simple effects analysis with Bonferroni correction revealed that, in the English-typical condition, participants scored significantly higher in the second block than in the first block ($p < .005$). In the Swedish-typical condition, block 2 scores were significantly higher than block 1 scores ($p < .001$), and block 4 scores were significantly higher than block 3 scores ($p < .01$). The analysis also revealed that scores were significantly higher in the English-typical condition in blocks 1 ($p < 0.01$), 2 ($p < 0.05$), and 3 ($p < 0.05$), but not block 4.

7. Discussion

As expected, participants in the low-endpoint condition showed better performance than participants in the high-endpoint condition (cf. Athanasopoulos & Bylund, 2013). Participants also showed a significant improvement across blocks,

indicating that they were successful in learning both the English-typical and Swedish-typical patterns. The significant effect of condition, however, shows that while participants began the experiment with an inclination to choose videos based on the English-typical pattern, they did not learn that pattern any faster than the Swedish-typical one. The fact that there is no significant difference between the conditions in block 4 suggests that, after training, participants were able to perform Swedish-typical categorizations nearly as accurately as English-typical ones.

Experiment 2: Supervised classification with verbal interference

8. Method

In an effort to determine the degree to which motion event classification is verbally mediated, experiment 2 was repeated with a verbal interference task.

8.1 Participants

The experiment was undertaken by 40 undergraduate psychology students (34 female, 6 male) from the University of Reading who were all monolingual native speakers of English and had not taken part in the previous experiment. Their ages ranged from 18 to 22.

8.2 Materials

The materials were identical to the ones used in experiment 1.

8.3 Procedure

The procedure was the same as in experiment 1, except that before each trial, the researcher read three randomly generated numbers between 11 and 99 to the participant. The participant was required to repeat these numbers back to the researcher throughout the trial, from just before the videos began to making a choice of A or B after they had viewed clip X. If the participant repeated a number or set of numbers incorrectly, they would be corrected and the trial would continue. After choosing A or B, the participant would stop repeating the numbers and the next set of

numbers would be read aloud to them. A different set of numbers was used for each of the 96 trials.

9. Results

Again, the 96 trials were split into blocks of 24 trials for the purposes of analysis, and the proportion of correct answers in each block was analysed.

The main effect of block approached significance, F (3, 111) = 2.69, p = 0.05. There was also a significant effect of condition, F (1, 37) = 6.87, $p < .05$. There was no significant interaction effect.
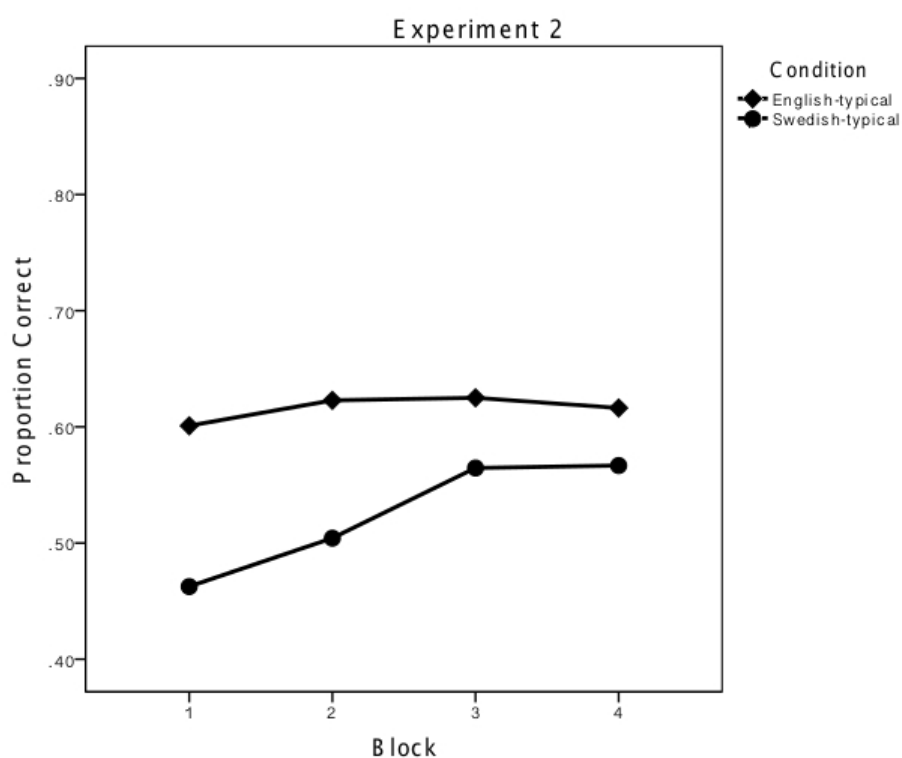

Figure 3. Results of motion video categorisation task in experiment 2. Participants in the English-typical condition scored higher across all blocks than those in the Swedish-typical condition. Significant improvements in categorization performance were only found in participants in the Swedish-typical condition.

A simple effects analysis with Bonferroni correction revealed no significant differences between blocks in the English-typical pattern, and significant differences between block 1 and blocks 3 and 4 ($p < .05$) in the Swedish-typical pattern, with block 2 showing no significant differences from any of the other blocks. Also shown

were significant differences between the two conditions in block 1 ($p < .001$) and block 2 ($p < .05$).

10. Discussion

As in Experiment 1, the (near) main effect of block indicates a significant improvement across blocks in the experiment, showing that participants were able to successfully learn the categorization patterns. However, the post-hoc analyses revealed that such improvement was most prominent in the Swedish-typical condition, and absent in the English-typical condition. The main effect of condition also shows that participants scored better on the low-endpoint pattern, as expected (cf. Athanasopoulos & Bylund, 2013).

It seems that with interference, the participants had a difficult time learning both patterns. However, as the simple effects analysis shows, their initial performance advantage on the English-typical pattern did not improve with training, and was erased by the end of the experiment, indicating that their categorization abilities had equalized.

11. General Discussion

There is an increasing amount of research showing that nonlinguistic and verbally mediated classification and perception processes are activated simultaneously, with language having an attenuating effect. Linguistic categories are spontaneously and automatically recruited in the process of making a categorical judgment. When the linguistic categories make a perceptual dimension more prominent by grammatically encoding it, then speed, accuracy, attention allocation, and category learning are facilitated, or enhanced. One of the hallmarks of this 'thinking with language' effect (Wolff & Holmes, 2011) is that verbal interference eliminates the language effect. If language is already engaged in a verbal task, it cannot be used as a facilitator in a concurrent non-verbal task.

The current paper set out to investigate three hypotheses related to the above observation. On one extreme the no verbal mediation hypothesis is placed, which predicts that event apprehension is completely impervious to linguistic influence.

Comparing overall tendencies between the two conditions regardless of the rate of learning, it seems that participants were always better at categorizing in the low endpoint condition, even under verbal interference. This replicates the finding in Athanasopoulos and Bylund (2013) and subsequent studies (Bylund, Athanasopoulos, & Oostendorp, 2013; Athanasopoulos et al., 2015) that perceptual similarity (in this case, the non-completion of motion in both the low endpoint and target alternates) is a strong cue for classification by similarity. However, this does not exclude attenuating effects of language as suggested by Athanasopoulos and Bylund (2013). Indeed, verbal interference clearly affected classification patterns since performance worsened in all conditions compared to the no interference experiment. This may not necessarily imply that verbal processes were dampened as it could also be attributed to a general effect of increased cognitive load that results from verbal interference (Baddeley, 2003). However, importantly, no learning occurred in the low-endpoint condition under verbal interference, in contrast to the no interference experiment. This suggests that participants must be relying on linguistic resources at least for those perceptual dimensions made salient in their native language.

The other extreme view, namely the maximal verbal mediation hypothesis, which predicts reliance on verbal codes throughout, regardless of whether the attended perceptual dimension is grammatically encoded in the observer's native language or not. However, the findings indicated a dissociation in the learning rate across the two learning conditions in the verbal interference experiment, such that learning continued to improve for the perceptual dimension not habitually encoded in the observers' native language. Could the encoding of the high-endpoint condition be verbal nonetheless? The answer is likely to be a negative one, because under the Working Memory view of encoding (Baddeley, 1991; 2003) under verbal interference: a) visual inputs should not be able to enter the phonological store (phonological coding is blocked); and (b) verbal rehearsal of the contents of the phonological store should be impossible. Since the verbal rehearsal system was already occupied in a secondary task, learning improvement in the high-endpoint condition is likely to have been facilitated by non-verbal means of encoding. Therefore the maximal verbal mediation hypothesis is not supported either.

The fact that learning was affected by verbal interference only in the condition that promoted the perceptual dimension associated with the observer's native

language suggests that individuals may recruit verbal processes online for the purposes of classification more readily when the stimuli to be classified are also habitually encoded in the native language. This was predicted by the selective verbal mediation hypothesis. What is the mechanism by which such selective effects of language obtain? The episodic buffer, the newest component of Baddeley's Working Memory model (Baddeley, 2000) that links Working Memory to Long Term Memory provides the likely mechanism for this effect to occur. One of the key roles of the episodic buffer is that it accesses long-term knowledge about language, grammar and the structure of sentences to bolster phonological short-term memory in the phonological loop. It is through this mechanism that we can access and utilise our stored language knowledge during ongoing memory and processing tasks. Our data support this key assumption in Baddeley's model because we show that when the phonological loop is engaged in a dual task (such as the number repetition task employed here) working memory is no longer able to benefit from long-term linguistic knowledge to aid learning.

Such selective effects of verbal interference are in line with findings from the language pathology literature, showing that in a patient with anomia, classification judgments of specific perceptual dimensions of stimuli such as colour, where reliance on language is paramount for the purposes of categorical judgments (Roberson, Davidoff & Braisby, 1999) was impaired, but classification judgments of colored objects based on taxonomic/thematic relationships not readily codable by linguistic labels was intact (Davidoff & Roberson, 2004).

Lupyan and Mirman (2013) found a similar pattern of dissociation in 12 patients with anomia. Participants were required to select all objects in an array that matched a specific criterion. In one condition, termed high-dimensional, it was possible to group objects on the basis of many different features (e.g. participants were told to 'select all the farm animals'). In the other condition, termed low-dimensional, grouping objects required attention to one specific feature whilst abstracting across other task-irrelevant dimensions (e.g. participants were told to 'select all the green objects'). Patients with anomia showed increased impairment in their ability to categorize low-dimensional relative to high-dimensional stimuli. The authors argued that linguistic processes facilitate performance on tasks utilising low-dimensional stimuli as these rely on more online support from language.

An earlier study (Lupyan, 2009) using verbal interference in normal populations yielded very similar patterns of selective effects: verbal interference disrupted categorization which was based on the ability to isolate perceptual dimensions readily amenable to verbalization, such as colour, but not categorization that required knowledge of thematic relationships between objects (e.g. that potato is the odd one out in the triad potato, balloon, and cake, since the latter two are linked by the theme of party).

The findings reviewed above, taken in conjunction with the current findings suggest more generally that when participants are used to relying on prior linguistic knowledge to make classification decisions, such as encoding features of stimuli like colours and in this case, the aspectual properties of motion events, the verbal mediation that would occur by default is disrupted by verbal interference. When however, the perceptual decisions involve stimuli that are not habitually coded through the verbal route, observers can, and do utilize non-linguistic means of classification. This conclusion is supported by the results of both experiments: Experiment 1 showed that speakers with no experience of an alternate categorization system are able to learn a new one quickly—by the end of the training session, participants were just as good at categorizing in the Swedish-typical pattern as they were with the English-typical pattern. Experiment 2, by introducing verbal interference, showed that learning to classify on perceptual dimensions already coded in the native language was disrupted, but learning to classify in a manner that was not already salient in the native language increased performance to a point that was statistically similar to the English-typical pattern.

12. Conclusion

To conclude, our results do not imply a change to perceptual mechanisms of classification since a) performance was better in the low-endpoint condition that is perceptually more similar to the intermediate-endpoint characteristic of the target stimulus, and b) verbal interference disrupted learning only in the condition where the perceptual dimension to be learned was also salient in the participant's native language, but not in the condition where it was not, suggesting that in the latter condition language-independent perceptual mechanisms are at play. Discussing the

interactive and dynamic nature of conceptual representation in the human brain, Lupyan (2012) warns that: "Viewing language as a part of an inherently interactive system with the capacity to augment processing in a range of non-linguistic tasks does not mean that performance on every task or representations of every concept are under linguistic control. Rather, the argument is that learning and using a system as ubiquitous as language has the potential to affect performance on a very wide range of tasks." (p. 3). Our results provide evidence to support Lupyan's cautious advice. Indeed, our data point to selective language influence depending on the associative or dissociative relationship between the linguistic features present in the observer's native language and the perceptual features of the stimuli at hand. When reliance on verbalization is removed, then classification is selectively affected for those stimuli that were more closely associated with linguistic structures in the observer's native language.

Interestingly, recent findings show that by selectively disrupting access to one or the other language bilingual individuals shift categorization behavior towards the patterns associated with the undisrupted language (Athanasopoulos et al., 2015). Such findings show not only the malleable nature of linguistically mediated cognition, but also that interpreting such transient effects of the kind found here as 'superficial' imposes a value judgment on an otherwise profound phenomenon: when observers are used to mediating their categorization decisions through language, they will do so in the majority of cases. This does not automatically preclude utilization of nonlinguistic mechanisms. Deconstructing the dichotomous question of whether language affects thinking or not and reformulating it as a question of degree, that is, to what extent does language rapidly and automatically affect seemingly non-linguistic perceptual processes (such as categorical perception, or event apprehension) has the potential to yield much more fruitful research endeavors and help clear out many of the misunderstandings that provide critics and supporters of the Whorfian hypothesis with erroneous argumentation in favour of one extreme or another.

References

Agrillo, C., & Roberson, D. (2009). Colour language and colour cognition: Brown and Lenneberg revisited. *Visual Cognition, 17*(3), 412-430. doi: 10.1080/13506280802049247

Athanasopoulos, P., & Bylund, E. (2013). Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish speakers. *Cognitive Science, 37*(2), 286–309. doi: 10.1111/cogs.12006

Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., Riches, N., & Thierry, G. (2015). Two languages, two minds: Flexible cognitive processing driven by language of operation. *Psychological Science.*

Baddeley, A. (1992). Working memory. *the interface between memory and cognition., 4*, 281-288.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423. doi: http://dx.doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*(3), 189–208. doi: 10.1016/s0021-9924(03)00019-4

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution.* Berkeley, CA: University of California Press.

Brown, R. W., & Lenneberg, E. H. (1954). A study in language and cognition. *The Journal of Abnormal and Social Psychology, 49*(3), 454-462. doi: 10.1037/h0057814

Bylund, E. & Athanasopoulos, P. (2014). Linguistic relativity in SLA: towards a new research programme. *Language Learning*, *64*, 952-985.

Bylund, E., Athanasopoulos, P., & Oostendorp, M. (2013). Motion event cognition and grammatical aspect: Evidence from Afrikaans *Linguistics* (Vol. 51, pp. 929).

Casasanto, D. (2008). Who's Afraid of the Big Bad Whorf? Crosslinguistic Differences in Temporal Language and Thought. *Language Learning, 58*, 63-79. doi: 10.1111/j.1467-9922.2008.00462.x

Casasanto, D., & Lupyan, G. (2015). All concepts are ad hoc concepts. In E. Margolis & S. Laurence (Eds.), *The Conceptual Mind: New Directions in the Study of Concepts* (pp. 543–566). Cambridge: MIT Press

Davidoff, J., & Roberson, D. (2004). Preserved thematic and impaired taxonomic categorisation: A case study. *Language and Cognitive Processes, 19*(1), 137-174. doi: 10.1080/01690960344000125

Flecken, M. (2011). Event conceptualization by early Dutch–German bilinguals: Insights from linguistic and eye-tracking data. *Bilingualism: Language and Cognition, 14*(1), 61–77. doi: 10.1017/s1366728910000027

Franklin, A., Clifford, A., Williamson, E., & Davies, I. (2005). Color term knowledge does not affect categorical perception of color in toddlers. *Journal of Experimental Child Psychology, 90*(2), 114-141. doi: http://dx.doi.org/10.1016/j.jecp.2004.10.001

Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition, 83*, 49–79.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences, 103*(2), 489-494. doi: 10.1073/pnas.0509868103

Goldstone, R. L., & Hendrickson, A. T. (2009). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(1), 69-78. doi: 10.1002/wcs.26

Harnad, S. R. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In S. R. Harnad (Ed.), *Categorial Perception: The Groundwork of Cognition*. Cambridge: Cambridge University Press

Imai, M., & Gentner, D. (1997). A crosslinguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition, 62*, 169–200.

Imai, M., & Mazuka, R. (2003). Re-evaluation of linguistic relativity: Language-specific categories and the role of universal ontological knowledge in the construal of individuation. In D. Gendter & S. Goldin-Meadow (Eds.), *Language in Mind: Advances in the Issues of Language and Thought* (pp. 430–464). Cambridge: MIT Press

Kay, P., & Kempton, W. (1984). What Is the Sapir-Whorf Hypothesis? *American Anthropologist, 86*(1), 65-79. doi: 10.1525/aa.1984.86.1.02a00050

Kersten, A. W., Meissner, C. A., Lechuga, J., Schwartz, B. L., Albrechtsen, J. S., & Iglesias, A. (2010). English speakers attend more strongly than Spanish speakers to manner of motion when classifying novel objects and events. *Journal of Experimental Psychology: General, 139*(4), 638–153. doi: 10.1037/a0020507

Langacker, R. W. (2008). *Cognitive Grammar: A Basic Introduction*. New York: Oxford University Press.

Lucy, J. A. (1992). *Grammatical Categories and Cognition: A Case Study of the Linguistic Relativity Hypothesis*. Cambridge: Cambridge University Press.

Lucy, J. A. (1997). The linguistics of "color". In C. L. Hardin & L. Maffi (Eds.), *Color Categories in Thought and Language*. Cambridge, England: Cambridge University Press

Lucy, J. A., & Gaskins, S. (2003). Interaction of language type and referent type in the development of nonverbal classification preferences. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in Mind: Advances in the Study of Language and Thought* (pp. 465–492). Cambridge, MA: The MIT Press

Lupyan, G. (2009). Extracommunicative functions of language: Verbal interference causes selective categorization impairments. *Psychonomic Bulletin & Review, 16*(4), 711–718. doi: 10.3758/PBR.16.4.711

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology, 3*. doi: 10.3389/fpsyg.2012.00054

Lupyan, G., & Mirman, D. (2013). Linking language and categorization: Evidence from aphasia. *Cortex, 49*(5), 1187-1194. doi: http://dx.doi.org/10.1016/j.cortex.2012.06.006

McWhorter, J. H. (2014). *The Language Hoax*. Oxford: Oxford University Press.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*(1), 39–57. doi: 10.1037/0096-3445.115.1.39

Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition, 108*(1), 155–184. doi: 10.1016/j.cognition.2008.02.007

Papafragou, A., Massey, C., & Gleitman, L. R. (2002). Shake, rattle, 'n' roll: The representation of motion in language and cognition. *Cognition, 84*, 189–219.

Papafragou, A., & Selimis, S. (2010). Event categorisation and language: A cross-linguistic study of motion. *Language and Cognitive Processes, 25*(2), 224–260. doi: 10.1080/01690960903017000

Pavlenko, A. (2014). *The Bilingual Mind: And What It Tells Us About Language and Thought*. Cambridge: Cambridge University Press.

Perry, L. K., & Lupyan, G. (2013). What the online manipulation of linguistic activity can tell us about language and thought. *Frontiers in Behavioral Neuroscience, 7*. doi: 10.3389/fnbeh.2013.00122

Pinker, S. (1994). *The Language Instinct: How the Mind Creates Language*. New York: William Morrow and Company.

Roberson, D., Davidoff, J., & Braisby, N. (1999). Similarity and categorisation: neuropsychological evidence for a dissociation in explicit categorisation tasks. *Cognition, 71*(1), 1-42. doi: http://dx.doi.org/10.1016/S0010-0277(99)00013-X

Rosch Heider, E. (1972). Probabilities, sampling, and ethnographic method: The case of Dani colour names. *Man, 7*(3), 448–466.

Talmy, L. (1991). *Path to realization: A typology of event conflation.* Paper presented at the Seventeenth Annual Meeting of the Berkeley Linguistics Soceity: General Session and Parasession on the Grammar of Event Structure, Berkeley.

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences USA*, *106*, 4567–4570.

Trueswell, J. C., & Papafragou, A. (2010). Perceiving and remembering events cross-linguistically: Evidence from dual-task paradigms. *Journal of Memory and Language, 63*(1), 64–82. doi: 10.1016/j.jml.2010.02.006

von Stutterheim, C., Andermann, M., Carroll, M., Flecken, M., & Schmiedtová, B. (2012). How grammaticized concepts shape event conceptualization in language production: Insights from linguistic analysis, eye tracking data, and memory performance. *Linguistics, 50*(4), 833–867. doi: 10.1515/ling-2012-0026

Whorf, B. L. (1956 [1940]). Science and linguistics *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. Cambridge, MA: MIT Press

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences, 104*(19). doi: 10.1073/pnas.0701644104

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(3), 253-265. doi: 10.1002/wcs.104