WHORF IN THE WILD:
NATURALISTIC EVIDENCE FROM HUMAN INTERACTION

Abstract

The past few decades have seen a full resurgence of the question of whether speakers of different languages think differently, also known as the Whorfian question. A characteristic of this neo-Whorfian enterprise is that the knowledge it has generated stems from psycholinguistic laboratory methods. As a consequence, our knowledge about how Whorfian effects play out in naturally occurring behaviour (i.e., 'in the wild') is severely limited. The current study argues that the time is ripe to redeem this evidentiary bias, and advocates a multidisciplinary approach towards the Whorfian question, in which insights from laboratory settings are combined with naturalistic data in order to yield a rounded picture of the influence of language on thought. To showcase the potential of such an approach, the study uses laboratory-generated knowledge on the influence of grammatical categories on cognition to interpret two examples of naturalistic human interaction and action in the domains of spatial navigation and scientific practice.

Introduction

Whorf is back. Big time. The classic question of whether the language we speak influences the way we think, typically associated with the work of Benjamin Lee Whorf, has experienced a remarkable resurgence over the past few decades. Indeed, since Steven Pinker (1994) wrote his so called 'obituary' of the Whorfian hypothesis in the early 1990s, hundreds of papers, tens of special issues, edited volumes, and monographs have offered an overwhelming amount of fascinating evidence lending support to the idea that the linguistic categories of our language may indeed point us to observing reality in particular ways (for recent overviews, see the collection of articles in Athanasopoulos, Bylund and Casasanto 2016a).[1] However, a common denominator of the demonstrations of Whorfian effects within this so-called 'neo-Whorfian' movement is that they occur mainly in experimental settings using the controlled paradigms afforded by cognitive psychology. Here, 'behaviour' or 'thought' is operationalized as some task that typically requires the participant to categorize or perceive a given set of stimuli in the laboratory (e.g., objects, colour, motion) on the premise that categorization is the cornerstone of human cognition (Harnad 1987). For example, Athanasopoulos, Dering, Wiggett, Kuipers and Thierry (2010) recorded brain responses in Greek and English speakers while presenting them with green and blue circles, and showed that Greek speakers perceived to a greater extent dark/light blue categorical contrasts than English speakers did, consistent with Greek colour terminology that encodes two categories of blue.

As with all lab-generated research, such tasks serve as proxies for the cognitive processes that underpin human behaviour outside the lab context. While these data may offer important insights into the architecture of the human mind, the fact remains that they are inherently artificial and, as a consequence, their ecological validity can be questioned (e.g., Hutchins 1995). An inevitable side effect is therefore that it is largely unknown how Whorfian effects play out "in the wild", that is, naturally occurring behaviour outside of the laboratory. This state-of-affairs has led several scholars to voice concerns about the ecological validity of our current knowledge about the influence of language on thought (e.g., Leavitt 2010; McWhorter 2014; Pavlenko 2014; Sidnell and Enfield 2012; Zinken 2012).

Against this background, the current paper takes a first systematic step towards redeeming this epistemological bias. The aim of the paper is to provide a conceptual argument for a multidisciplinary approach to linguistic relativity. To achieve this, we first bring in naturalistic examples of human interaction and action in the form of illustrative vignettes, using as an interpretive lens the knowledge gained from neo-Whorfian laboratory

studies, and second, we outline a programme for ecologically valid Whorfian research that incrementally moves the field forward by placing multidisciplinarity at its heart. The standpoint taken in the paper is that a combination of experimental and naturalistic data is likely to yield deeper insights into the effects of language on thought than either may yield in isolation.

The paper is organised in the following way: We start by defining the scope of our study and the nomenclature used therein. Next, we present current findings on motion event construal, which then serves as the backdrop for our analyses of two naturalistic examples of Whorfian effects in human interaction and action. Finally, we provide a programmatic outline for an ecologically valid approach to language and thought.

Whorf in the wild: Defining the terms of engagement

The scope of the paper is circumscribed to engage with the issue of ecological validity without entering into the several controversies that surround the Whorfian question (for critical discussions see Pinker 1994; Casasanto 2008; McWhorter 2014). This means that, first, we will not engage in any discussion as to whether the Whorfian principle is true or false. Such binary interpretation is not only counter-productive, but also obsolete in view of the growing body of evidence showing that the effects of language on thought are infinitely more complex than that (Wolff and Holmes, 2010). Second, we do not claim to hold the truth about what Whorf meant in his various formulations of his principle, or how he envisaged its empirical realisation, if he did at all. Such claims tend to carry disciplinary bias, may lead to strawman argumentation, and have likely caused the controversy associated with the principle since its inception (see discussions in Bylund and Athanasopoulos 2014a, Athanasopoulos, Bylund and Casasanto 2016b; Casasanto, 2008; Fishman, 1960). Instead, we will assume the commonly held view among linguists, anthropologists, and psychologists (e.g., Casasanto 2008; Duranti 2012; Levinson 2012; Lucy 1992; Regier and Kay 2009; for critical discussion, see Jarvis 2016; Pavlenko 2016) that the essence of the Whorfian principle is that our understanding and interpretation of the world is influenced by the linguistic categories of the language(s) we speak, such that our thoughts and actions are mediated by the lexical and grammatical structures made available in our language(s). Put another way, our language affects our behaviour, in predictable ways.

What we need to define, however, is what we mean by 'in the wild'. Following Hutchins (1995), we use the phrase to refer to "naturally occurring culturally constituted human activity" (p. xiv). This use is thus different from notions such as "pensée sauvage"

(Levi-Strauss 1962) and domesticated mind (Goody and Goody 1977), and instead capitalises on the distinction between the laboratory where Whorfian effects are studied in isolation, and the everyday world, where Whorfian effects play out in habitual behaviour (Lucy 1992). It is in this context therefore important not to equate experiments in the field with behaviour in the wild. Whereas the former typically entails applying experimental paradigms in the participants' natural environment, the latter refers to uninhibited and non-elicited behaviour (independently of where such behaviour occurs).

A primitive empirical demonstration of behaviour in the wild was in fact recounted by the scholar who coined the term linguistic relativity in the first place, namely Benjamin Lee Whorf. Whorf developed the hypothesis by observing behaviour in the wild: industrial accidents occurring as a result of people smoking and dropping lit cigarette butts near gasoline drums labelled '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 casual nonverbal behaviour (Whorf 1956b).

A more recent example comes from Steve Levinson and his study of absolute vs relative spatial orientation in speakers of Guugu Yimithirr (GY), an Australian Aboriginal language, and Dutch (Levinson 1997; see also Lucy 1992 for brief anecdotes from the domain of object categorisation). Dutch and most other Standard Average European speakers tend to give and understand directions in an egocentric, relative way: "After the traffic lights, take the first right, then the second left, and then you'll see the super market on your left." But speakers of GY (and many other Polynesian, Native American, and African languages) tend to use a system of absolute orientation that relies on fixed geographic directions: "After the traffic lights, drive south, and then on the second crossing drive west, and you'll see the super market directly to the north." One of the most notable anecdotes that Levinson (1997) recounts concerns him having difficulty understanding directions given by a GY speaker about where to find fish in the local store:

"They have plenty of fish fillet in the store."
"I've never seen it; where?"
"On this side" [gesturing] "in the frozen food container, far end this side."

Levinson reports that he and his GY interlocutor were standing in a hospital 45 kilometres from the said store. Levinson was expected to observe that the gesture of his interlocutor was to the northeast, so that next time he was in the store he would look in the northeast corner.

This naturalistic interaction suggests that habitual ways of talking about space have consequences for understanding location and direction during talk-in-interaction and behaviour in the wild.

Motion events in language and cognition

We have chosen to showcase naturalistic Whorfian effects in the domain of motion, given its centrality in human everyday life. The encoding of motion in language and cognition is a major area of investigation within neo-Whorfian research in general (e.g., Pederson and Bohnemeyer 2010), and more recently applied linguistics (e.g., Bylund and Athanasopoulos 2015a). It has been demonstrated that speakers of different languages attend differently to the goal, or endpoint, of motion. For instance, imagine two people walking along a road at the end of which there is a church. If you speak a language that encodes the aspectual distinction of imperfectivity (and its subcategories, e.g., the progressive[2]) on an obligatory scale (labelled 'aspect languages' in this literature, e.g., English, Algerian Arabic, Russian, Spanish), you are prone to 'zoom in' on the motion itself and the path along which it occurs, whereas if you are a speaker of a language that lacks this grammatical distinction (i.e., 'non-aspect languages', e.g., Afrikaans, German, Swedish) you are more likely to zoom out' and include the motion endpoint in your construal (Athanasopoulos and Bylund 2013; Bylund 2008, 2011; Bylund & Athanasopoulos 2014b; Bylund, Athanasopoulos and Oostendorp 2013; Flecken, von Stutterheim and Carroll 2014; von Stutterheim and Nüse 2003; von Stutterheim et al. 2012). This is presumably because speakers of aspect languages encode the ongoing phase of events on a habitual basis (through obligatory marking of imperfective aspect), thereby excluding its endpoints, whereas speakers of non-aspect languages instead construe events from a holistic perspective.

This zooming in and out of temporal frames has been operationalised in a number of different ways in the lab. For the elicitation of verbal behaviour, speakers are typically asked to describe a set of specially designed videos showing motion towards an endpoint/goal, but the video ends before the endpoint is reached. Such retelling tasks may also be combined with online methods such as eye-tracking. Findings from these paradigms show that speakers of aspect languages are less likely to mention and look at the (potential) endpoint of a given motion event than speakers of non-aspect languages (e.g., von Stutterheim et al. 2012). For the elicitation of non-verbal behaviour (i.e., outside of the speech act), a commonly used technique is the so-called triads-matching (Lucy 1992), or ABX task, where the participant is shown a stimulus triad: video A, which depicts motion with a reached endpoint ("high

endpoint alternate"); video B, which depicts motion without any discernible endpoint ("low endpoint alternate"); and video X, which depicts motion with a potential (but not reached) endpoint ("intermediate"). This is a forced-choice task, and the participant has to decide whether video A or B is most similar to video X (the order of videos A and B is counterbalanced). Findings here show that speakers of aspect languages tend to judge the low endpoint alternate video as more similar to X, compared to speakers of non-aspect languages who show the reverse pattern (e.g., Athanasopoulos and Bylund 2013; Athanasopoulos et al. 2015; Bylund & Athanasopoulos 2014b, 2015b).

Naturalistic Examples of Whorfian Effects in the Domain of Motion

We use two examples of naturalistic interaction and action, presented as vignettes, to illustrate how cross-linguistic differences in the conceptualisation of motion influences everyday behaviour and scientific thinking. Example 1 illustrates a communication breakdown arising from language-specific interpretations of directions. Example 2 provides a demonstration of how researchers' native languages can bias experimental design in the very same domain, namely cross-linguistic differences in motion cognition.

Example 1: At the Duke's Estate

Sociocultural approaches to first language (L1) and second language (L2) development (Kramsch, 2004; Lantolf and Thorne, 2007; Lucy and Wertsch, 1987; Vygotsky, 1987) consider language as one of the key mediators of human mental functioning and a primary instrument of socialization and acculturation. Therefore, such approaches have the potential to readily expand the scope of the Whorfian endeavour to a more ecologically valid empirical realm by investigating the effects of language on social interaction. An empirical example of how this can be achieved is provided by Sidnell and Enfield (2012). The authors focused on how epistemic authority (defined as "the action of agreeing with what someone has just said while simultaneously signalling that one has greater authority to have said it" (p. 304) is asserted in talk-in-interaction across different sociocultural contexts. By analysing the assertion of epistemic authority in spontaneous conversations across a number of languages, they show that the conversational devices used for this purpose lead, in some linguistic contexts (e.g. Caribbean English Creole) to topic dropping, and in others (e.g. Finnish) to topic expanding.[3] The finding is novel and striking: different ways of actualizing the same communicative goal shape the ongoing interaction in different directions. Sidnell and Enfield's (2012) approach focuses on the effects of conversational devices on talk-in-

interaction. However, the potential is there to study talk-in-interaction effects of grammatical categories beyond just the verbal level (as proposed by Duranti 2012 and Lucy 2016), as the following example illustrates.

Example and analysis[4]

Several years ago, we were engaged in a collaborative research project the aim of which was to ferret out whether crosslinguistic differences in grammatical aspect encoding influences thinking about motion events beyond speech production. The reported incident occurred while we were working on this project. It took place during a walk in the English countryside, following a footpath connecting a town with a small village. The walk proceeds through a private Estate owned by an English Duke, who has allowed access to the footpath for the public. There is no other way to get from the town to the village on foot except than to go along the public footpath that cuts through the Duke's Estate. The footpath ends just outside the final approach to the village, upon which one must cross a large field to reach the village. The first village building visible on that side is the village church and cemetery (see Figure 1). However, it is not uncommon for people to walk towards another footpath parallel to the large field and try to reach the village through a gated entrance. In doing so, they would be unknowingly trespassing on private property, namely the Duke's Estate. There are no prominent signs to warn people off. Instead, unaware trespassers are addressed via an intercom system controlled by a security guard who can also receive video images through CCTV. Oblivious trespassing of this kind must occur frequently, because the advice given is almost always the same formulaic instruction (or some slight variant thereof): "To get out of the Duke's estate, walk across the field diagonally". This level of granularity to describe the required motion in the scenery depicted in Figure 1 is apparently enough for English speaking interlocutors to understand what it is they need to do. But when a multilingual group of friends (including the authors of this paper) decided to take this walk, this set of instructions caused confusion for some of them.

Specifically, two of our friends, a Spaniard and a German, had walked ahead of the rest of the group, and had made the classic trespassing error that any newcomer to the village makes, described above. As we observed this from a distance, we rushed to warn them to take a different path. As soon as we reached them, we heard the familiar instructions over the intercom. The exchange proceeded as follows:

"To get out of the Duke's estate, walk across the field diagonally"

The Spaniard promptly begins to walk, while the German remains in situ, appearing hesitant (perhaps confused). Then the Spaniard points to the path across the field (at the end of which there is a church, see Figure 1). After a few moments of careful visual examination and consideration of the scene the German says:

"Ah, so you mean we should walk towards the church?"

With some delay, the Spaniard replies:

"Er… yes"

[FIGURE 1 HERE]

Two rich points[5] can be identified in this exchange, concerning the endpoint of the action (the church) vs. the action itself (walk across the field diagonally). Our Spanish friend seemed perfectly content with the English speaking guard's instructions, and his subsequent action indicated that he understood them without hesitation. Our German friend, however, did not take immediate action, and this presents the first rich point in the interaction. Now clearly, since she is a very advanced speaker of English, this cannot have been because she failed to understand the semantic content of the instructions. Her body language, looking at the field and trying to work out how to map what she heard with the affordances the visual scene presented her with, indicated that at a pragmatic level she was assuming that the guard was being cooperative, and she was trying to be cooperative herself: Indeed, her request for clarification in the form of an attempt to recast the instructions indexed her willingness to follow the guard's expressed intention. The problem, in more precise terms, was one of mutual manifestness (Sperber and Wilson, 1995), defined as follows: "A fact is manifest to an individual at a given time if and only if he is capable of representing it mentally and accepting its representation as true or probably true" (Sperber and Wilson, 1995: 39). Mutual manifestness is a product of the moment and thus distinct from mutual knowledge between speaker and addressee: no prior assumptions about the context of the exchange are required for interpretation. Rather, the addressee constructs these assumptions based "on what he can perceive in his immediate physical environment, or on the basis of assumptions already stored in memory" (Pilkington, 2000: 62).

As we mentioned earlier, Spanish and English are aspect languages that zoom in on the ongoingness of an event, excluding its endpoints, whereas German is a non-aspect language, with a holistic event construal that includes endpoints.[6] So one logical assumption that follows is that for the German speaking addressee in our example above to be able to mentally represent the security guard's instructions, she had to inspect the immediate physical environment and reconstruct the directions in a way that conformed to assumptions already stored in memory, namely the tendency in her native language to structure information involving motion in a way that includes the goals or endpoints of said motion. Visually searching for an endpoint in this naturalistic setting represents a real-life instance of a phenomenon observed in speakers of non-aspect languages in the laboratory with eye-tracking. Similarly, the Spaniard's immediate shift of attention to the path of the suggested action mirrors the immediate allocation of attention to motion paths in speakers of aspect languages (Flecken et al., 2015).

Recasting the same directions in the form of a goal-directed action (walk towards the church) may have enabled the German speaker to confirm, or manifest, the veracity of the directions, but it also led to the second rich point in the exchange, where the Spanish speaker initially hesitated in his confirmation of his German friend's statement. The pause in-between occurred because our Spaniard friend literally looked up into the horizon to verify that there was indeed a church at the end of the path described by the security guard. Again, this seems to not only align with the definition of mutual manifestness and its manifestation in the first rich point, but also points to typological differences in conceptualization and information structure between Spanish and English on the one hand, and German on the other, as the likely cause of the rich points. Indeed, the Spaniard's native language, as well as the language of interaction, are both aspect languages and laboratory findings show that the use of multiple aspect languages exacerbates the defocusing of motion endpoints (Bylund and Athanasopoulos 2014b).

To assess the accuracy of our interpretation that these communication near-breakdowns (and subsequent repairs) were indeed caused by differences in endpoint focus (and not some hitherto unknown cultural difference between speakers and addressees, or any other contextual variable unaccounted for), we ran our analysis by the interlocutors, as means of member checking. The Spaniard confirmed that he indeed had not noticed the church until pointed to it by the German, but instead had understood that by simply following the path they would find their way out. The German, in contrast, pointed out that to her the instructions had seemed to lack a clear sense of direction since it was not mentioned *whereto*

they were supposed to be heading. As a result, she first hesitated and then made the connection with the church as the implicit 'whereto'. Indeed, experimental findings show that German speakers tend to construe a potential endpoint even when one is not visibly present (von Stutterheim 2003). Both speakers also confirmed that they had had no problems in understanding the words and constructions in the guard's instructions.

Implications

The exchange described above provides an illustrative example of the way in which grammatical categories can influence behaviour during talk-in-interaction. Had the exchange been studied without the knowledge of crosslinguistic differences in endpoint encoding, it would have been challenging to grasp the breakdown and repair in all their richness. It could of course have been hypothesised that these events occurred because the speakers seemed to differ in their goal-orientation, as this is a possible reading of the exchange even without insights into neo-Whorfian research on endpoint encoding. However, in that case it would still be difficult to pinpoint what underlies such differences.

Adopting the well-documented laboratory effects of grammatical aspect on motion event cognition as an interpretative frame of the exchange instead offers several advantages. Firstly, it has a firm empirical ground in numerous independent studies confirming linguistic structure as the main cause of endpoint encoding behaviour. Second, it offers a nuanced understanding of the nature of this behaviour, showing that while behaviour in the lab is easily malleable by experimentally manipulating the language context of operation (Athanasopoulos et al. 2015), behaviour in the wild may not always be as malleable. Specifically, even though the German interlocutor was operating in an aspect language context (English), and the conversational context involved two speakers of aspect languages (the security guard, and her Spanish friend), her interpretation of directions was still ostensibly influenced by her native language (for further evidence of L1-induced patterns of event construal in L2 speakers, see Athanasopoulos, Damjanovic, Burnand and Bylund 2015, Bylund and Athanasopoulos 2015b; Donoso and Bylund 2014; Flecken, Carroll, Weimar and von Stutterheim 2015). Indeed, as she reported to us, the situation she found herself in induced a certain degree of stress (having trespassed and trying to understand the guard's directions). This, in turn, may have prompted a more L1-derived endpoint orientation, since it

is well-documented that affective neural pathways are more strongly connected to the L1 than to the L2 (Harris et al., 2003).

Alternative explanations of the breakdown and repair may take into account cognitive sex-differences, namely that the female brain excels in verbal tasks whereas the male brain is better adapted to visual-spatial tasks (e.g. Joseph, 2000). Such explanation could be compatible with the first part of the exchange: one could attribute the German speaker's (female) confusion to a combination of less emphasis on visuospatial traits in favour of enhanced verbal skills, which are more likely to trigger reliance on language-specific cognitive structures. Similarly, one could attribute the Spanish speaker's (male) quick understanding of the guard's instructions as typical of his enhanced visuospatial skills. But such an account would leave half the data unexplained. The second part of the exchange, and our subsequent member checking, confirmed that the German speaker emphasised visuospatial traits of the scene (the church-endpoint) not attended to by the Spanish speaker. Hence alleged cognitive sex differences in enhanced visuospatial skills cannot adequately account for the data. While we cannot claim that the Whorfian explanation is the only plausible explanation (no doubt many other personal history variables may be at play), it nonetheless remains the least implausible explanation for the communication near-breakdowns observed here, and is in fact compatible with socio-cultural evidence that, all other things being equal, there are minimal differences in verbalisation susceptibility between men and women (Cameron 2009). We believe that without knowledge of the typological differences between German and Spanish, and the rich laboratory evidence showing differences in how German and Spanish speakers verbalise and cognitively process events, interpretation of such cross-cultural exchanges would be left to endless speculation and debate. Essentialistic interpretations are thus more easily avoided when the linguistic and cognitive complexity of multilingual behaviour is known.

In summary, the example illustrates a unique instance of how crosslinguistic cognitive differences may shape the ongoing behaviour and interaction between interlocutors. To the best of our knowledge, links between language and cognition at this level of specificity are seldom made when observing behaviour in the wild.

Example 2: In the Laboratory
Background
The event in Example 1 suggests Whorfian effects in the wild, actualised in the moment of communication. However, behaviour in the wild also entails actions that are realised over a

longer time-frame, such as actions and behaviour arising from collaborative activities towards a common goal (Hutchins 1995). Example 2 illustrates this, focusing on the domain of scientific practice.

The role that language plays in the scientific enterprise is a classic topic that has been discussed for centuries -- or even millennia -- in a variety of different disciplines (e.g., Bennett and Hacker 2003; Bourdieu 2004; Harris 1980). A famous example is the debate between physicists Niels Bohr and Albert Einstein over quantum mechanics, which Bohr himself interpreted as resulting from divergent understandings of certain concepts, leading him to contend that "[w]e are suspended in language in such a way that we cannot say what is up and what is down" (cited in Petersen 1963: 10). As in the Bohr-Einstein case, a major part of the debate and inquiry on language and science has concerned whether terminology induces preconceived ideas about the phenomenon it refers to, and thus whether scientific thought can ever be put into words without becoming 'tainted'. Actually, an example of this may be found in the trajectory of nascent empirical approaches to linguistic relativity. Leavitt (2011), in his historical treatise of the Whorfian idea complex, made the point that once what Whorf referred to as the 'principle' of linguistic relativity was turned into a 'hypothesis' (cf. Brown and Lenneberg 1953) there was a shift in the evidentiary basis towards hypothesis-testing paradigms emanating from the cognitive psychology laboratory.

Whorf's approach to the role of language in science sought to move beyond nomenclature alone, attempting to understand how crosslinguistic differences in categorisation influenced the scientific enterprise. This idea was later echoed by Fishman (1960), and potential examples of such instances are found in the study of colour perception where allegedly universal colour categories were modelled on English colour terms (e.g., Berlin and Kay, 1969) (for further discussion, see Pavlenko 2014; Wierzbicka, 2008). So far, however, few attempts have been made to pinpoint the role that language-specific categorisation patterns plays in the actual research process itself. The event described below constitutes one such example.

Example and analysis

When we started collaborating on our motion event project several years ago, we decided to use a triads-matching task, given its widespread use in neo-Whorfian research. As described in the section 'motion events in language and cognition', this is a similarity judgement task where the participant has to decide whether a target event X is more similar to alternate A or B. The stimuli had been made available to us by Christiane von Stutterheim's team at

Heidelberg University. A crucial step in the design was to classify them into three different categories depending on their endpoint orientation, as per Table 1.

[TABLE 1]

The process of categorising the clips into these different categories is as central as it is sensitive: inappropriate classifications may produce biases or inconsistencies in the stimulus set, which, in the end, will yield unreliable experimental data. Having selected a subset of 47 clips (based on their common characteristics, e.g., number of participants, motion type, and scenery) out of a pool of over a 100, we proceeded to classify these clips according to Table 1. Given the centrality of this step in the design process, as well as practical issues (e.g., working in different countries), we decided to first do the classifications independently of one another, and then compare our lists. To our surprise, our lists were quite different and it turned out rather difficult to agree how the clips should actually be classified. We have assessed the inter-rater reliability of our classifications. For the subset of clips as a whole, our agreement reached 72.3%, which is just below the commonly recommended 75%. As indicated by Cohen's kappa, this agreement was 'moderate', $\kappa = .59$. A closer look at our classifications reveals that the disagreements varied as a function of endpoint orientation category. For candidates to the category "high degree of endpoint orientation", 81.8% ($\kappa = .67$) of our classifications aligned. In contrast, the classifications of candidates to the "intermediate degree" and "low degree" categories showed poor conformity, only 55.6% ($\kappa = .05$) and 47.8% ($\kappa = .05$), respectively, suggesting that agreement was at chance level. The low incidence of conformity here arose from the fact that clips that the first author (P.A.) tended to classify as "intermediate degree", were classified as "low degree" by the second author (E.B.).

[FIGURE 2A, B, C HERE]

An example of a clip that triggered heated debate amongst us is shown in stills in Figure 2A, 2B, 2C. Here, E.B. would argue that, beyond any doubt, the house is clearly visible, and the fact that it is at the end of the path along which the women are walking makes it a given candidate for the endpoint of the motion: as such, the clip should be categorised as "intermediate degree". This endpoint-based interpretation of the clip corresponds to a holistic viewing frame, in which the agents in motion (the women) and the endpoint (the house)

(Figure 3A) are included. P.A., on the other hand, would argue that even though the house might be visible, the clip depicted two women out for a walk, and to interpret the house as being the endpoint of the motion was a leap of imagination: therefore, the clip should be categorised as "low degree". This interpretation with emphasis on ongoingness, in turn, is consistent with an immediate viewing frame, zooming in on the motion and excluding the endpoint (Figure 3B).

[FIGURE 3A, B, C]

As disagreements and exchanges of this type arose throughout the classification process it started to dawn upon us that they may actually be reflections of the very thing we sought to test with our experiment: the effects of language on thought. There are several factors providing support for this interpretation. During this phase of the research preparation process, E.B. was working in Sweden and, while being proficient in the aspect languages English and Spanish, he used his native non-aspect language Swedish as primary means of communication.[7] Swedish has no grammatical category to mark imperfective aspect. P.A., on the other hand, who is a native speaker of Greek[8], had by then been studying and working in the UK for over 10 years. P.A.'s primary languages of communication, English and Greek, were languages that obligatorily mark imperfective aspect. In other words, the degree to which our classifications differed was largely isomorphic to the crosslinguistic differences that the experiment yielded later on. Moreover, our strong agreement on the high-endpoint alternates is perfectly aligned with robust findings that the crossing of a boundary (into a house, out of a house) is perceptually salient and may override language-induced categorisation biases (Athanasopoulos & Bylund 2013; Regier and Zheng 2007; Slobin 1996).

Even though, before carrying out the classifications, we had made sure that the three categories of endpoint orientation were carefully defined, the classification bias may have started already at the time the different categories were created, such that, for instance, the label 'medium degree of endpoint orientation' invoked different pictures in our minds as to what an exemplar of this category should look like. The bias was later exacerbated by our different conceptualisations of the clips themselves[9].

It could possibly be argued, though, that these differences in endpoint classification preferences were autosuggested, stemming from a desire to believe that grammatical aspect

does influence cognition - a fair share of subjective beliefs is indeed found in the history of linguistic relativity research (Pinker 1994; Pullum 1988) - and then behaving accordingly. However, there are reasons to believe this was not the case. When embarking on the research project, we had both felt very sceptical as to whether crosslinguistic differences in endpoint descriptions would really influence higher order cognitive processing, since contemporary research on other types of motion events (focusing on path and manner) had reported minimal - if any - crosslinguistic differences (e.g., Papafragou et al. 2002). Thus, obtaining a null result in this new area of motion cognition (goal-orientation and grammatical aspect) would have been perfectly in line with previous findings, providing compelling evidence that cognitive processing in the general motion domain is subject to universal-perceptual principles rather than linguistic categories.[10]

Implications

The event described constitutes a seldom documented, detailed example of the effects of linguistic categories (i.e., beyond pure terminology) on scientific practice.

The example further highlights two additional aspects of this problem. First, it shows that simply having knowledge of another language does not create immunity against linguistic bias in scientific practice. Even though E.B. was a proficient L2 speaker of two aspect languages, his classifications were still very consistent with the linguistic context in which they had been carried out. It is of course possible that a monolingual speaker of Swedish could have exhibited an even stronger endpoint-preference, but studies show that language context of operation is a strong determinant of behaviour in the laboratory as well, even in advanced L2 speakers (Athanasopoulos et al. 2015; Bylund and Athanasopoulos 2017). This example illustrates that L2 knowledge, in and of itself, is not a prophylactic against native language induced biases. This resonates with Pavlenko's (2014) belief that making bilingualism an academic requirement would not automatically eliminate the risks that academics still rely on their native or dominant language in their research endeavour.

However, something that may actually reduce such risks is working with multilingual and multicultural research teams that can bring multicompetent perspectives (Cook and Li Wei 2016). The advantage of such research teams has been highlighted in the context of ethnographic research (e.g., Creese, Takhi and Blackledge 2016; Martin 2012), but the constellation is yet to be recognised as a standard methodological requirement in linguistic relativity research. The consequences of monolingual teams are potentially far-reaching: Had the team members in the example above shared native language, it would have increased the

risk of biasing the selection of videoclips, which ultimately could have skewed the results in unpredictable ways. Instead, in the current case, each member brought his own language-specific perspective of the meaning of endpoint-orientation to the research task at hand, and then, through a process of negotiation, the final set of experimental stimuli was co-constructed[11].

Another important aspect of Example II is the reflexive relationship it illustrates: the influence of language on thought *in research on* the influence of language on thought (see also Silverstein 1979). The attested effect should however not come as a surprise. If language-specific categories may indeed influence our thinking, why should scientific thinking be exempt from this influence? To think that it is, would be consistent with the controversial view of the scientific mind as detached, obeying only the laws of logic, immune to any experiential factors (such as language background, ethnicity, gender, etc.) (for further discussion see the correspondence between Kay and Kuehni (2008) and Wierzbicka (2008)).

Towards a Wilder Whorf
The vignettes presented above serve to illustrate the potential of bringing together different data types for gaining a fuller picture of the influence of language on thought. We believe that the development of this enterprise toward a workable research programme will be best served by a closer disciplinary alliance than currently exists between lab-generated research and ethnographic methods of recording behaviour in the wild. In essence, we argue that the optimal way forward is multidisciplinarity. There is a wealth of studies within linguistic anthropology positing that behavioural patterns are upheld through, or mediated by, linguistic patterns, drawing on long-term solid ethnographic evidence (e.g., Duranti 1994; Hanks 1991). In the enterprise we envisage, such evidence is combined with controlled experiments that allow the researcher to establish a direct link (or even causality) between linguistic categories and behaviour.

More specifically, we propose a tripartite evidentiary cycle towards a more nuanced and rounded understanding of Whorfian effects, as per Figure 4, based on the following three components: 1) Typology, broadly defined as language usage patterns, including (but not limited to) lexicogrammatical categories, information structure (i.e. what the speaker chooses to encode, and how they phrase it), and conversational devices; 2) the Lab refers to behaviour elicited through controlled experiments; 3) the Wild refers to natural behaviour occurring in natural surroundings. These components are interlinked in distinct ways: Typology may generate hypotheses to be tested both in the lab and in the wild. Evidence from the wild may

also be (re)interpreted against typology, as our examples illustrate (hence the bidirectionality of arrow B), but this is unlikely to be the case for experimental evidence, as experiments are designed to test hypotheses derived from typology (hence the unidirectionality of arrow C). A bidirectional relationship also exists between the wild and the lab (arrow A). Crucially, there need not be a hierarchical or chronological relationship between naturalistic and laboratory data. The one data type is neither necessarily superior to the other, nor must it be obtained before the other. Sometimes observations of naturalistic behaviour precedes - and triggers - laboratory studies (e.g., Levinson 1997), other times naturalistic behaviour may be interpreted or reinterpreted through knowledge gained from the lab (as in Example 1). Ideally, one type would corroborate the other. If the ethnographer observes cultural practices that she interprets as stemming from specific linguistic devices, and then finds that lab studies also show the effects of such linguistic devices in controlled experimentation, then she may point to that evidence as indeed substantiating her interpretation. Conversely, if the psycholinguist's experiment yields a finding that shows the effect of a specific grammatical category on behaviour, and then finds that ethnographic work attributes similar behaviours in the wild to the same grammatical category, the psycholinguist can then point to that evidence as strengthening their lab finding. Equally useful would be instances where behaviour from the lab and from the wild are discordant, because this generates new impetus for the researcher to investigate the root of the discordancy and shed further light on factors affecting the behaviour under study.

[FIGURE 4]

We are aware that the practicalities of a multidisciplinary approach may seem daunting, as it would require a researcher who is trained in both linguistic ethnography and experimental methods, and a sufficient amount of observation and long-term participation in the community being observed. However, similar to the case made for multilingual research teams earlier in the paper, we suggest that multidisciplinary research teams, comprised by for instance linguists, anthropologists, and psychologists, would hold the greatest of promise for undertaking the endeavour outlined in figure 4. Building up such teams and methodological expertise may be a lengthy process, but one that we believe is achievable, if providing a more rounded picture of the Whorfian question is the aim.

There are however smaller steps that can be taken in the interim. If we now imagine that arrow A is a continuum representing human behaviour, one end of which is the wild, the

other end being the lab, then we can generate fertile empirical ground for methodological practices that are situated at various points within this continuum. The evidentiary harvest may reveal aspects of behaviour that corroborate or reveal something novel that could not have been observed by adhering to strict disciplinary methodological principles. Below we offer suggestions towards this end.

Anecdotes from the lab: An easy way to start is for psycholinguists to report on participants' behaviour that is not the specific targeted dependent variable of the experiment, but still provides relevant information. For instance, in our own experiments on motion event description (where the dependent variable is 'endpoint mentioned/omitted), we have noticed that speakers of non-aspect languages remark that the descriptions are difficult to carry out because it is not always obvious whereto people are going. Such anecdotal evidence would serve as corroborating evidence, but could also potentially reveal hitherto unknown aspects of behaviour that could serve as pivots for further research (see also Lucy 2004 for similar anecdotal reports during his experiments).

Take the lab to the wild: This step entails designing experimental tasks that are in fact suited to behaviour in the wild in the first place. An example of this is Levinson's (1997) extension of the spatial cognition research, where participants were asked to follow and interpret directions outdoors. In addition, owing to fast technological developments, modern day psycholinguistic equipment is more mobile than it used to be, and thus less constrained to a specific physical space. An obvious way to take the lab to the wild is then to make use of instruments that record attention allocation (i.e., eye-tracking), galvanic skin response, heart rate, etc. These techniques are already used 'in the wild' in other fields (e.g., consumer behaviour research, Gidlöf, Wallin, Dewhurst and Holmqvist 2013), suggesting there is potential utilising these techniques in Whorfian research. All the above suggestions of course entail that the observer is a conscious participant of the study, but they would record or elicit behaviour in a naturalistic environment without artificial or proxy representation of reality constrained by the visual affordances of a lab computer screen.

Take the wild to the lab: In the lab, because experimental control is key, tasks are often designed to tap a specific cognitive or perceptual process. However, in real life, as becomes evident in the examples, relativistic effects may manifest themselves in a range of different mental operations, all occurring at the same time or in close connection. Our accumulated knowledge about relativistic effects is now such, at least in some perceptual domains, that we can move beyond the traditional categorisation tasks and start looking at more interactive tasks inside the laboratory, using more ecologically valid stimuli where

possible. For example, Pavlenko et al. (2017) elicited verbal descriptions of real paintings containing various hues and shades of blue in Russian speakers through semi-structured interviews, without using the standardized colour charts (e.g., Munsell) used typically in colour research. If we wanted to extend this research to the wild, we could envisage a non-verbal colour task that would for example ask participants to colour in a picture using a set of paints, the availability of which is manipulated in such a way that tests the salience of the colour categories of interest in the painter's mind. In a similar vein, immersive virtual reality laboratory paradigms can be developed to emulate, as much as possible, naturalistic settings (see Casasanto & Jasmin 2017).

Concluding remarks

The aim of this paper was to systematically engage with the Whorfian hypothesis in a more ecologically valid way. To this end, we first used insights generated from neo-Whorfian laboratory studies to analyse two examples of Whorfian behaviour in the wild, and then proceeded to articulate a programmatic outline for future research endeavours. The vignettes presented in the paper clearly illustrate the potential of different data types for gaining a fuller picture of the influence of language on thought. Indeed, as pointed out in the introduction, it was by observing such real life examples of (mis)understanding that inspired Whorf himself to come up with the linguistic relativity hypothesis. Whorf likened the ubiquitous nature of effects of language on thinking to the unconscious nature of our background experience informing our worldview:

"…if a rule has absolutely no exceptions, it is not recognized as a rule or as anything else; it is then part of the background of experience of which we tend to remain unconscious. Never having experienced anything in contrast to it, we cannot isolate it and formulate it as a rule until we so enlarge our experience and expand our base of reference that we encounter an interruption of its regularity. The situation is somewhat analogous to that of not missing the water till the well runs dry, or not realizing that we need air till we are choking." (Whorf, 1956, p. 209).

Interpreting a scene or a social action through the lens of language is an unconscious process. The intersubjectivity issues and subsequent repairs described in both of our examples can be precisely viewed as instances of this 'interruption of regularity' that Whorf describes in the above extract. We can create and control this interruption of regularity in the psycholinguist's

laboratory, or observe and analyse it using the ethnographer's toolkit. As the phenomenon studied is the same, there seems to be no reason why a more holistic methodological approach should not be adopted.

Our examples demonstrate that it is indeed possible to show effects of linguistic categories beyond just the laboratory, documenting behaviours in the wild that may be rooted in the intricately diverse grammatical and lexical systems that typically form the basis de rigueur of traditional Whorfian and neo-Whorfian approaches. We propose that a multidisciplinary approach would offer the much needed missing link between the neo-Whorfian movement and naturalistic behaviour that has eluded Whorfian studies to date. In our experience, the topic of linguistic relativity, or language and thought in general, tends to spark considerable interest among students, laypeople, and journalists. However, when people hear about the methodologies used in neo-Whorfian research, their interest either fades because of the apparent artificial nature of the experiments, or it sparks unwarranted cross-cultural generalisations that lack the firm basis in linguistic typology and the nuance provided by experimental methods. In the absence of naturalistic data, our knowledge about Whorfian effects outside of the laboratory will remain speculative, at best. Likewise, as long as experimental data are lacking, our knowledge about the effects of a given linguistic structure on cognitive processing will remain crude.

References

Agar, M. 2006. 'Culture: Can You Take It Anywhere?' *International Journal of Qualitative Methods*, 5 (2) https://sites.ualberta.ca/~iiqm/backissues/5_2/PDF/agar.pdf   retrieved 23/03/2018.

Athanasopoulos, P. and **E. Bylund**. 2013. 'Does Grammatical Aspect Affect Motion Event Cognition? A Cross-Linguistic Comparison of English and Swedish Speakers.' *Cognitive Science* 37 (2): 286–309. https://doi.org/10.1111/cogs.12006.

Athanasopoulos, P., E. Bylund and **D. Casasanto**. 2016a. 'Introduction to the Special Issue: New and Interdisciplinary Approaches to Linguistic Relativity.' *Language Learning* 66 (3): 482–86. https://doi.org/10.1111/lang.12196.

Athanasopoulos, P., E. Bylund and **D. Casasanto**. 2016b. 'New and Interdisciplinary Approaches to Linguistic Relativity (Special Issue).' *Language Learning* 66 (3). https://doi.org/10.1111/lang.12196.

Athanasopoulos, P., E. Bylund, G. Montero-Melis, L. Damjanovic, A. Schartner, A. Kibbe, N. Riches, and **G. Thierry**. 2015. 'Two Languages, Two Minds: Flexible Cognitive Processing Driven by Language of Operation.' *Psychological Science*, 26 (4), 518–26. https://doi.org/10.1177/0956797614567509.

Athanasopoulos, P., L. Damjanovic, J. Burnand, and **E. Bylund**. 2015. 'Learning to Think in a Second Language: Effects of Proficiency and Length of Exposure in English Learners of German.' *The Modern Language Journal* 99 (S1): 138–53. https://doi.org/10.1111/j.1540-4781.2015.12183.x.

Athanasopoulos, P., B. Dering, A. Wiggett, J. Kuipers, and **G. Thierry**. 2010. 'Perceptual Shift in Bilingualism: Brain Potentials Reveal Plasticity in Pre-Attentive Colour Perception.' *Cognition* 116 (3): 437–43. https://doi.org/10.1016/j.cognition.2010.05.016.

Bennett, M. and **P. Hacker.** 2003. *Philosophical Foundations of Neuroscience*. Wiley.

Berlin, B. and **P. Kay.** 1969. *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California Press.

Bylund, E. 2008. 'Procesos de conceptualización de eventos en español y en sueco: diferencias translinguísticas [Conceptualization processes of events in Spanish and Swedish: crosslinguistic differences]' *Revue Romane* 43 (1): 1–24. https://doi.org/10.1075/rro.43.1.02byl.

Bylund, E. 2009. 'Effects of Age of L2 Acquisition on L1 Event Conceptualization Patterns.' *Bilingualism: Language and Cognition* 12 (3): 305–322. https://doi.org/10.1017/S1366728909990137.

Bylund, E. 2011. 'Language-Specific Patterns in Event Conceptualization: Insights from Bilingualism.' In *Thinking and Speaking in Two Languages*, edited by A. Pavlenko. Bristol: Multilingual Matters, pp. 108–142.

Bylund, E. and **P. Athanasopoulos.** 2014a. 'Linguistic Relativity in SLA: Toward a New Research Program.' *Language Learning* 64 (4): 952–85. https://doi.org/10.1111/lang.12080.

Bylund, E. and **P. Athanasopoulos.** 2014b. 'Language and Thought in a Multilingual Context: The Case of IsiXhosa.' *Bilingualism: Language and Cognition* 17: 431–43.

Bylund, E. and **P. Athanasopoulos.** 2015a. 'Introduction: Cognition, Motion Events, and SLA.' *The Modern Language Journal* 99 (S1): 1–13. https://doi.org/10.1111/j.1540-4781.2015.12175.x.

Bylund, E. and **P. Athanasopoulos.** 2015b. 'Televised Whorf: Cognitive Restructuring in Advanced Foreign Language Learners as a Function of Audiovisual Media Exposure.' *The Modern Language Journal* 99 (S1): 123–37. https://doi.org/10.1111/j.1540-4781.2015.12182.x.

Bylund, E., P. Athanasopoulos and **M. Oostendorp.** 2013. 'Motion Event Cognition and Grammatical Aspect: Evidence from Afrikaans.' *Linguistics* 51: 929–55.

Bylund, E. and **T. Dick.** 2019. 'Pioneer: Benjamin Lee Whorf.' In *SAGE Research Methods Foundations*, edited by P. Atkinson, S. Delamont, M. Hardy, and M. Williams. New York: SAGE. doi: 10.4135/9781526421036852591.

Cameron, D. 2009. 'Sex/Gender, Language and the New Biologism'. *Applied Linguistics* 31/2: 173-192.

Casasanto, D. 2008. 'Who's Afraid of the Big Bad Whorf? Crosslinguistic Differences in Temporal Language and Thought'. *Language Learning* 58/s1: 63–79.

Casasanto, D. and **K. Jasmin.** 2017. 'Virtual Reality' in De Groot, A. M. B. and P. Hagoort (eds): *Research Methods in Psycholinguistics and the Neurobiology of Language*. Wiley, pp. 174–189.

Cook, V. and **Li Wei.** 2016. *The Cambridge Handbook of Linguistic Multi-Competence*. Cambridge University Press.

Creese, A., J. K. Takhi and **A. Blackledge.** 2016. 'Reflexivity in Team Ethnography'. In Martin-Jones, M. and D. Martin (eds): *Researching Multilingualism: Critical and Ethnographic Perspectives*. Routledge, pp. 203-214.

Donoso, A. and **E. Bylund.** 2014. "The Construal of Goal-Oriented Motion Events by Swedish Speakers of L2 Spanish." In *The Acquisition of Spanish as a Second Language from Different First Languages*, edited by J. Tiffany and S. Perpiñán. Amsterdam: John Benjamins, pp. 234–254.

Duranti, A. 2012. 'Comment on Sidnell, Jack, and N. J. Enfield. 'Language Diversity and Social Action: A Third Locus of Linguistic Relativity''. *Current Anthropology* 53/3: 321–322.

Flecken, M., M. Carroll, K. Weimar and **C. von Stutterheim.** 2015. 'Driving Along the Road or Heading for the Village? Conceptual Differences Underlying Motion Event Encoding in French, German, and French–German L2 Users.' *The Modern Language Journal* 99 (S1): 100–122. https://doi.org/10.1111/j.1540-4781.2015.12181.x.

Flecken, M., C. von Stutterheim and **M. Carroll.** 2014. 'Grammatical Aspect Influences Motion Event Perception: Findings from a Cross- Linguistic Non-Verbal Recognition Task.' *Language and Cognition* 6 (01): 45–78. https://doi.org/10.1017/langcog.2013.2.

Gidlöf, K., A. Wallin, R. Dewhurst, and **K. Holmqvist.** 2013. 'Using Eye Tracking to Trace a Cognitive Process'. *Journal of Eye Movement Research* 6/1: 1-14.

Hanks, W. 1991. *Referential Practice: Language and Lived Space Among the Maya*. University of Chicago Press.

Harnad, S. 1987. 'Psychophysical and Cognitive Aspects of Categorical Perception: A Critical Overview'. In Harnad, S. (ed.): *Categorical Perception: The Groundwork of Cognition*. Cambridge University Press, pp. 1-52. http://cogprints.org/1571/.

Harris, R. 1980. *The Language-Makers*. Duckworth.

Harris, C., A. Aycicegi, and **J. Gleason.** 2003. Taboo words and reprimands elicit greater autonomic reactivity in a first than in a second language. *Applied Psycholinguistics*, 4: 561–78.

Hutchins, E. 1995. *Cognition in the Wild*. MIT Press.

Jarvis, S. 2016. 'Clarifying the Scope of Conceptual Transfer'. *Language Learning* 66/3: 608–635. https://doi.org/10.1111/lang.12154

Joseph, R. 2000. 'The evolution of sex differences in language, sexuality and visual-spatial skills,' *Archives of Sexual Behavior* 29/1: 35–66.

Kay, P. and **R. Kuehni**. 2008. 'Correspondence'. *Journal of the Royal Anthropological Institute* 14/4: 886–87. https://doi.org/10.1111/j.1467-9655.2008.00536_1.x.

Kramsch, C. 2004. 'Language, Thought, and Culture'. In Davies, A. and C. Elder (eds): *The Handbook of Applied Linguistics*. Blackwell, pp. 235–61. http://onlinelibrary.wiley.com/doi/10.1002/9780470757000.ch9/summary.

Lantolf, J. and **Thorne, S.** 2007. 'Sociocultural Theory and Second Language Learning'. In VanPatten, B. and J. Williams (eds): *Theories in Second Language Acquisition*. Lawrence Erlbaum, pp. 201-224.

Levinson, S. 1997. 'Language and Cognition: The Cognitive Consequences of Spatial Description in Guugu Yimithirr'. *Journal of Linguistic Anthropology* 7/1: 98–131. https://doi.org/10.1525/jlin.1997.7.1.98.

Levinson, S. 2012. 'Foreword'. In Levinson, S C. and P. Lee (eds): *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. MIT Press, pp. vii–xv.

Leavitt, J. 2010. *Linguistic Relativities: Language Diversity and Modern Thought*. Cambridge University Press.

Lévi-Strauss, C. 1962. *La pensée sauvage*. Plon.

Lucy, J. 1992. *Language Diversity and Thought*. Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511620843

Lucy, J. 2004. 'Language, culture, and mind'. In Achard, M. and Kemmer, S. (eds.): *Language, Culture, and Mind*. University of Chicago Press, pp. 1-21.

Lucy, J. 2016. 'Recent advances in the study of linguistic relativity in historical context'. Language Learning 66: 487-515.

Lucy, J. and **J. V. Wertsch.** 1987. 'Vygotsky and Whorf'. In Hickmann, M. (ed.): *Social and Functional Approaches to Language and Thought*. Academic Press, pp. 67-86.

Martin, D. 2012. 'A Critical Linguistic Ethnographic Approach to Language Disabilities in Multilingual Families'. In Gardner, S. and Marilyn Martin-Jones (eds): *Multilingualism, Discourse, and Ethnography*. Routledge, pp. 305–18.

McWhorter, J. 2014. *The Language Hoax*. Oxford University Press.

Papafragou, A., C. Massey, and **L. Gleitman.** 2002. 'Shake, Rattle, "n" Roll: The Representation of Motion in Language and Cognition'. *Cognition* 84/2: 189–219. https://doi.org/16/S0010-0277(02)00046-X.

Pavlenko, A. 2014. *The Bilingual Mind*. Cambridge University Press.

Pavlenko, A. 2016. 'Whorf's Lost Argument: Multilingual Awareness.' *Language Learning* 66/3: 581–607. https://doi.org/10.1111/lang.12185

Pavlenko, A., S. Jarvis, S. Melnyk, and **A. Sorokina**. 2017. 'Communicative Relevance: Color References in Bilingual and Trilingual Speakers'. *Bilingualism: Language and Cognition* 20/4: 853–866. https://doi.org/10.1017/S1366728916000535.

Pederson, E. and **J. Bohnemeyer**, eds. 2010. *Event Representation in Language and Cognition*. Cambridge, UK: Cambridge University Press. http://www.cambridge.org/se/knowledge/isbn/item2710292/?site_locale=sv_SE.

Petersen, A. 1963. 'The Philosophy of Niels Bohr'. *Bulletin of the Atomic Scientist* 19: 8–14.

Pilkington, A. 2000. *Poetic Effects: A Relevance Theory Perspective*. John Benjamins.

Pinker, S. 1994. *The Language Instinct: How the Mind Creates Language*. Penguin.

Pullum, G. 1988. *The Great Eskimo Vocabulary Hoax and Other Irreverent Essays on Language*. Equinox.

Sidnell, J. and **N. Enfield.** 2012. 'Language Diversity and Social Action: A Third Locus of Linguistic Relativity'. *Current Anthropology* 53/3: 302–33. https://doi.org/10.1086/665697.

Silverstein, M. 1979. 'Language Structure and Linguistic Ideology' 193-247. In P. Clyne, W. Hanks, and C. Hofbauer (eds.): *The Elements. A Parasession on Linguistic Units and Levels*. Chicago Linguistic Society, pp. 193-247.

Sperber, D. and **D. Wilson**. 1995. *Relevance: Communication and Cognition* (2nd edition). Blackwell.

von Stutterheim, C., M. Andermann, M. Carroll, M. Flecken, and **B. Schmiedtová**. 2012. 'How Grammaticized Concepts Shape Event Conceptualization in Language Production'. *Linguistics* 50/4: 833–67. https://doi.org/10.1515/ling-2012-0026.

Wierzbicka, A. 2008. 'Why There Are No "Colour Universals" in Language and Thought'. *Journal of the Royal Anthropological Institute* 14/2: 407–25. https://doi.org/10.1111/j.1467-9655.2008.00509.x.

Wolff, P. and **K. Holmes**. 2011. 'Linguistic Relativity'. *Wiley Interdisciplinary Reviews: Cognitive Science* 2/3: 253–265. https://doi.org/10.1002/wcs.104

Whorf, B. L. 1956. *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. MIT Press.

---

[1] Bibliometric figures (Bylund and Dick 2019) indeed show that the citation frequency of Whorf (1956) has sky-rocketed over the past decade (86 citations per year in the 1980s; 140 citations per year in the 1990s; 277 citations per year in the 2000s; 420 citations per year in the 2010s).

[2] In essence, imperfective aspect denotes a 'from-within' perspective of a given situation (e.g., *Mary is playing the piano*), encoding the unfolding phase of the situation without attention to the situation's beginning or ending (see Binnick 1990, Comrie 1976, Dahl 1985).

[3] As explained by Sidnell and Enfield (2012), in Finnish, free word order permits the expressions *se on* ('it is') and *on se* ('is it') to be used to assert independent epistemic access to the same fact (*se on*), or agreement to a previously asserted observation or statement while also signalling a different/additional perspective (*on se*). However, the latter construction is also used in contexts where the speaker wishes to elaborate her point (aside from asserting epistemic authority). Sidnell and Enfield report a conversation where K says to L that the Autumn weather in the countryside is beautiful and L agrees using *on se*. L then proceeds to describe how she is visiting her countryside cottage the following week for the last time because she is selling it. In Caribbean English Creole, epistemic authority may instead be signalled by prefacing a repeat of a prior speaker's utterance in the form of a question with "if." For instance, a conversation about a naughty child among three adults, one of whom is the child's grandmother but the others more distantly related, the conversation will usually end by the grandmother replying to the others' assertions that the child is rude by asserting the naughtiness of the child by if-prefacing: *if he's rude?* However, an if-prefacing question is also a common conversational device used to initiate repair of a prior turn that is expressed as a polar (yes-no) question (K: *If I am here for the what?* L: *The festival*. K: *Yes*) and typically functions as a turn closer. (Contrary to Finnish, where use of a specific word order leads to topic expanding, in Caribbean English Creole, use of the relevant construction leads to topic dropping. In the example involving the grandmother of the rude child, after epistemic agreement has been expressed with an if-prefaced question, the topic of the conversation among the adults completely changes.

[4] The personal and geographic details of the example have been modified to ensure anonymity.

[5] We take Agar's definition of rich point (Agar, 2006): "Rich points are those surprises, those departures from an outsider's expectations that signal a difference between LC1 and LC2" (2006: 2). LC1 and LC2 stand for Languaculture 1 and Languaculture 2. Languaculture is a term coined by Agar to indicate his belief that language knowledge includes cultural knowledge. A later post on his personal blog defines a rich point as "meaning a difference based on experience that indexes a major cultural difference worthy of attention as a research focus" (http://www.ethknoworks.com/blog.htm?post=715651)

[6] Spanish has an obligatory distinction in the past tense between perfective and imperfective aspect, and a semi-optional distinction between progressive aspect and non-progressive aspect in the present and the past tenses. English has an obligatory distinction between progressive and non-progressive aspect in the present and the past tenses (Comrie 1976, Dahl 1985).

[7] Swedish lacks grammatical categories to denote imperfective aspect (Dahl 1985). In order to express imperfectivity, lexical circumlocutions or body posture verbs may be used. These constructions do not, however, seem to occur with motion verbs (Bylund 2008, 2009).

[8] Greek has an obligatory distinction in the past between perfective and imperfective aspect (Dahl 1985).

[9]The bias pervades to this day, in many instances of our collaboration. P.A. took the picture that is displayed in figure 1. When E.B. saw it, he immediately commented that the path is foregrounded too much, and the church (endpoint) isn't as prominent as it should be.

[10] In fact, some iterations of the experiment manipulating working memory load yielded evidence for universal-perceptual principles in event cognition.

[11] Specifically, this was done through a process of joint re-analysis whereby the controversial clips were discussed and the reasons for their different classifications weighed until consensus was reached. In cases where consensus could not be reached, clips were discarded, so as to rule out any remaining bias.