Short title/running header: Bilingual time perception

Full title: Cognitive restructuring: Psychophysical measurement of time perception in bilinguals*

Names and affiliations of the authors:
Panos Athanasopoulos, Department of Linguistics and English Language, Lancaster University; Department of General Linguistics, Stellenbosch University.
Emanuel Bylund, Department of General Linguistics, Stellenbosch University; Centre for Research on Bilingualism, Stockholm University.

*Acknowledgments: The authors acknowledge funding from the Bank of Sweden Tercentenary Foundation. We are grateful to Daniel Casasanto for sharing technical details of his experimental protocol, and providing feedback on an early draft of this paper. We are also grateful to two anonymous reviewers for their constructive feedback, and to Aina Casaponsa and Robyn Berghoff for statistical advice.

Address for correspondence:
Professor Panos Athanasopoulos
Department of Linguistics and English Language
County South
Lancaster University
Lancaster
United Kingdom
LA1 4YL

Email address: p.athanasopoulos@lancaster.ac.uk

Orchid ID: https://orcid.org/0000-0003-1601-0476

Cognitive restructuring: Psychophysical measurement of time perception in bilinguals

Abstract

This paper explores the link between the metaphoric structure TIME IS SPACE and time perception in bilinguals. While there appear to be fundamental commonalities in the way humans perceive and experience time regardless of language background, language-specific spatiotemporal metaphors can give rise to differences between populations, under certain conditions. Little is known, however, about how bilinguals experience time, and the specific factors that may modulate bilingual temporal processing. Here, we address this gap by examining L1 Spanish - L2 Swedish bilinguals in a psychophysical task. Results show that duration estimation of dynamic spatial configurations analogous to L2-specific temporal metaphors is modulated by L2 proficiency. In contrast, duration estimation of spatial configurations analogous to the L1 metaphorical expressions appears to be modulated by the age of L2 acquisition. These findings are discussed in terms of associative learning and cognitive restructuring in the bilingual mind.

Keywords: spatiotemporal metaphors, duration estimation, cognitive restructuring

Introduction

How long is a minute? Or can a minute even be long, or short, for that matter? It depends. According to some languages, a minute is big or small, but not long. Would these linguistic conventions have any bearing on our experience of the passage of time? Recently, it has been shown that this may indeed be the case. When asked to estimate the duration of dynamic events, speakers of different languages are systematically misled by different types of task-irrelevant information: Native speakers of languages where time is usually talked about as if it were long or short (e.g., English and Swedish), are likely to be distracted by visual cues relating to distance, thus perceiving physically stretched out lines as also being stretched out in time. In contrast, speakers of languages where time is mainly depicted as big or small (e.g., Greek and Spanish), are instead misled by visual cues relating to quantity or volume (Bylund & Athanasopoulos, 2017; Casasanto, Boroditsky, Phillips, Greene, Goswami, Bocanegra-Thiel, Santiago-Diaz, Fotakopolou, Pita & Gil, 2004).

That language is a potentially powerful factor in shaping our perception of the passage of time should come as no surprise, since various aspects of temporal cognition (e.g., the sequencing of events, see Boroditsky, 2011) have been shown to be influenced by linguistic categories. What kind of linguistic experience is then necessary to bring about such an effect? If the way one's native language expresses time meddles with time perception, does the learning of a new way of talking about time bring about the same effect? Short term changes in time perception have been documented in laboratory training paradigms (e.g. Casasanto, 2005). However, the long-term effects of learning a new way to talk about duration, as in the case of bilingualism, have only recently begun to attract empirical interest. Specifically, research shows that bilingual individuals may flexibly switch between different temporal behaviours, depending on the language they operate in (Bylund & Athanasopoulos, 2017). While this suggests that bilingual time perception is a promising venue for testing the interaction of language and cognition in the human mind, little is still known about the extent to which bilinguals actually converge with monolingual speakers in their time perception patterns, and crucially, which particular aspects of the bilingual experience may shape this outcome.

The aim of the current study is to address this gap by testing time perception in second language (L2) users. Specifically, using an established psychophysical task of duration reproduction, we examine the degree to which L2 users of Swedish with Spanish L1 resemble native speakers of either language, against several background variables known to shape cognitive restructuring in bilinguals (see Bylund & Athanasopoulos, 2014a).

Background

Time in language and mind

Starting with Whorf's (1956) contested analysis of temporal concepts in the Hopi language and their alleged impact on thought, temporal cognition has been a classic testbed in linguistic relativity research (the idea that language affects thinking, in predictable ways). The backdrop of this line of research is the insight that temporal language is typically built on spatial frames of reference (e.g., H. H. Clark, 1973; Traugott, 1978). While we cannot see or touch time, we often talk about it as if we could. Consider the following expressions:

1a) *The table is behind us*

1b) *The meeting is behind us*

2a) *The table was pushed past the window*

2b) *The meeting was pushed past the deadline*

3a) *A long table*

3b) *A long meeting*

Temporal expressions of this kind are examples par excellence of conceptual metaphors, whereby concepts from a more tangible domain (in this case space) are used to denote abstract phenomena (in this case time) (Lakoff & Johnson, 1980). Systematic studies reveal that space is indeed central for the linguistic construal of time, as spatial metaphors may be applied to the core distinctions of deictic (D) time, sequence (S) time, and temporal (T) span, as identified by Núnez and Cooperrider (2013). D-time and S-time have in common that they refer to the order or sequence of a given set of events (as in 1b & 2b), but they differ in their anchoring of events, with D-time assuming a temporal landmark (the deictic now), and S-time instead linking the events between themselves. T-span, which is the focus of this paper, is different from these concepts in that it relates to the duration or temporal magnitude of an event (as in 3b), with no reference to any other temporal entity. Whereas there is an extensive body of research on the spatial building blocks of talking and thinking about D- and S-time (Boroditsky, 2011), the concept of T-span is to date much less investigated.

A central feature about linguistic expressions of T-span is that they often reference spatial magnitude. Crucially, as mentioned above, these magnitudes may be of different natures across languages. In Germanic languages, there is a pronounced tendency to talk about duration as a distance. Consider the following English examples:

4a) *It's been* <u>*long*</u> *since we last saw each other*

4b) *We'll be in touch* <u>*shortly*</u>

In contrast, in Greek and some Romance languages, duration is preferentially expressed as volume or quantity. Here, the distance-based spatial metaphors underlined in (4) would instead be quantity-based, as in the following Spanish examples:

5a) *Ha pasado <u>mucho</u> tiempo desde la última vez que nos vimos*

(approx. 'It's been <u>much</u> time since the last time we met')

5b) *Hablamos dentro de <u>poco</u>*

(approx. 'We'll talk within <u>little'</u>)

It is nonetheless possible to use quantity-based metaphors to talk about time in Germanic languages, as in *She had spent so <u>much</u> time on preparing the memorandum*. Notably, however, expressions like these do not necessarily relate to duration, but rather give a connotation of iterativity (i.e. she had worked on the memorandum over and over). Likewise, in for instance Spanish, it is possible to talk about time in terms of distance, as in *No habían visto la costa por un tiempo muy largo* (approx. 'they hadn't seen the coastline for a very long time'). Here, the use of a distance-based metaphor conveys a particular stylistic value of emphasis (i.e., they hadn't seen the coastline for ages). These usages are, in other words, exceptions, as shown by Casasanto et al. (2004) and Bylund and Athanasopoulos (2017), and the overwhelming pattern for speakers of English and Swedish is to express duration in terms of distance, and for speakers of Greek and Spanish to preferentially use metaphors of quantity.

The question then arises as to whether these routinized linguistic expressions exert any influence on T-span perception? The two studies conducted to date suggest that this is indeed the case. Using a psychophysical paradigm of duration reproduction, Bylund and Athanasopoulos (2017) and Casasanto et al. (2004) asked participants to reproduce the duration of lines extending across the screen, or boxes/containers filling up with liquid. These animations had been manipulated such that two lines could have the same duration (e.g., 3 seconds) but have different physical lengths (e.g., 200 pixels or 400 pixels), the rationale being that in the participant's mind spatial information would interfere with temporal information, creating the sense of physically longer lines as being longer in time. Results showed that this kind of spatial interference varied as a function of language, in that speakers of English and Swedish tended to perceive physically longer lines as also being stretched out in time, whereas this was not the case for the Greek and Spanish speakers (who instead exhibited spatial interference when reproducing the container filling up duration).

Casasanto (2008) attributed such spatial interference patterns to associative learning: when people use a linguistic metaphor for time, they activate the corresponding mental metaphor. As people use the dominant and less-dominant metaphors in their language, they activate one mental metaphor more frequently than the other(s). In doing so, they would strengthen this particular associative mapping. Interestingly, however, this postulation can be qualified in two important ways.

Firstly, the language-specific spatial interference effect was found for stimuli that were harder to process. The paradigm utilises 9 different line growths/container fill levels, and 9 different durations, fully crossed to yield 81 experimental stimuli. Bylund and Athanasopoulos (2017) divided their stimuli into extreme, consisting of the two shortest and two longest spatial displacements and durations, and medium stimuli (consisting of the five spatial displacements and durations in-between the extreme ones). An independent task established that the extreme stimuli were easier to process than the medium stimuli. The fact that Casasanto et al. (2004) and Bylund and Athanasopoulos (2017) found language effects only for medium length/duration stimuli that were harder to estimate duration for, points to the role of language as a cognitive tool to solve difficult or complex tasks. This is compatible with findings from other domains such as colour, where cross-linguistic differences were obtained for hard-to-discriminate but not for easy-to-discriminate colour stimuli (Winawer, Witthoft, Frank, Wu, Wade & Boroditsky, 2007) and in line with more general theoretical accounts of Whorfian effects such as the language-as-strategy-hypothesis proposed by Gennari, Sloman, Malt and Fitch (2002) and Papafragou and Selimis (2010): the more difficult or cognitively demanding the task becomes, the more language is recruited to help with the task at hand.

Secondly, effects of language-specific metaphors on duration estimation obtain exclusively in verbal contexts, that is, when the stimuli are preceded by the word for duration in the participant’s native language. In a version of the experiment that used visual instead of verbal prompts, Bylund and Athanasopoulos (2017) found no cross-linguistic differences between Swedish and Spanish monolinguals, in contrast to the experiment where the verbal prompts were present, suggesting that some kind of online top-down feedback from language (Lupyan, 2012) must be the primary mechanism driving the cross-linguistic differences in duration perception. This interpretation received further empirical support in Bylund and Athanasopoulos (2017) who also tested Swedish-Spanish bilinguals, varying the language context of experiment instructions and the verbal prompts in order to see whether this would yield distinct language-specific patterns of time perception. In this modified experiment, the bilinguals, who were Swedish foreign-language users of Spanish, turned out to reproduce time differently depending on the language of the prompts present in the experiment. Specifically, when participants received Spanish prompts they were more susceptible to spatial interference in the filling-containers than in the growing lines condition, but when receiving Swedish prompts, they were more susceptible to spatial interference in the growing lines condition than in the filling containers condition.

To interpret these findings, Bylund and Athanasopoulos (2017) complemented the general associative learning mechanism proposed by Casasanto (2008) with a framework of predictive processing (e.g., A. Clark, 2013; Hohwy, 2013; Kanai, Komura, Shipp & Friston, 2015; Lupyan & Clark, 2015). Under this account, the construction of a percept is viewed as a continuous and concurrent interplay between top-down predictions and bottom-up sensory signals. The top-down predictions are expectations based on prior knowledge (e.g., encyclopaedic and contextual knowledge), but, compatibly with associative learning, are also constantly updated by sensory input. Language may function both as a source of prior knowledge, as well as a bottom-up contextual cue. The weighting of downward-flowing predictions and upward-flowing sensory signals is modulated, among other things, by the ambiguity of the input: a stimulus that is hard to discriminate is more likely to trigger greater recruitment of prior knowledge, than a stimulus that is easy to discriminate. This may then explain why the medium-long/-large stimuli, which were indeed harder to reproduce, exhibited language-specific interference. Moreover, the effects of verbal prompts on duration reproduction documented in both the monolingual and bilingual speakers are consistent with the view of language as a powerful bottom-up contextual cue, whereby words activate semantic networks, which then function as a knowledge prior. The question, from a bilingual development perspective, is whether the second/newly acquired language can ever exert a powerful enough influence to meddle with already established time representations, and what factors drive this development. The next section turns to focus on precisely these questions.

Language and thought in bilinguals

The question of whether learning a second language meddles with thought patterns was discussed already by von Humboldt (1836/1963) and Whorf (1956), who in their classic writings on language and thought speculated as to whether learning a new language entails learning a new way of observing reality. Von Humboldt expressed two extreme points in what scholars like Pavlenko (1999), Cook (2002) and Athanasopoulos (2011) construe as a continuum of cognitive restructuring. At one end, the shift or restructuring view entails complete internalisation of L2-specific concepts: “The learning of a foreign language should mean the gaining of a new standpoint toward one's world-view" (von Humboldt 1836/1963, p. 294). 'Learn a new language, get a new soul', as a famous proverb says. At the other end resides the idea of conceptual entrenchment: "If it is not always purely felt as such, the reason is one so frequently projects one's own world-view, in fact one's own speech habits, onto a foreign language" (von Humboldt, ibid). Echoes of this idea can be found in the writings of modern scholars such as Slobin (1996) who speculated that cognitive patterns already established in the L1 are exceptionally resistant to restructuring later in life. In a less hardline tone, Lucy (2004; 2016) has discussed the idea of semantic or conceptual accents. Just like an individual may carry the speech accent of their native language when they speak a second, so may that individual carry the semantic accent of their native language concepts.

However, long-standing speculation aside, it was only recently that bilingualism started entering the agenda of mainstream empirical linguistic relativity research (for a recent synopsis see Athanasopoulos, Bylund and Casasanto, 2016). Central questions here concern the extent to which bilinguals may flexibly switch between language-specific thought patterns, the degree to which bilinguals resemble monolingual speakers of their respective languages, and the factors that underlie this degree of (non)resemblance.

Extant evidence on cognitive flexibility indicates that bilingual individuals may switch between distinct cognitive behaviours as a function of the linguistic context. For instance, Athanasopoulos, Bylund, Montero-Melis, Damjanovic, Schartner, Kibbe, Riches and Thierry (2015a) found that depending on the language of instruction of the experiment, as well as the language elements present during the experiment (in the form of verbal interference), German-English bilinguals would categorize motion either according to German-specific patterns or English-specific patterns (see also Kersten, Meissner, Lechuga, Schwartz, Albrechtsen & Iglesias, 2010). In the domain of time specifically, it has been found that bilingual individuals may indeed switch between different temporal representations. Miles, Tan, Noble, Lumsden and Macrae (2011) found that culture-specific cues lead Mandarin-English bilinguals to conceptualise the flow of time as either vertical (i.e., Mandarin-like) or horizontal (i.e., English-like). However, a similar change of behaviour as a function of culture-specific cues was also reported in monolinguals, reflecting the fact that Mandarin utilises BOTH a vertical and horizontal temporal representation (Yang & Sun, 2016). As mentioned earlier, Bylund and Athanasopoulos (2017) found that Spanish-Swedish bilinguals would reproduce duration differently depending on the language of prompts of the experiment.

There are however other studies that have failed to show effects of language context on bilingual cognition. For motion categorisation and recognition memory, Filipovic (2011) found that the behaviour of English-Spanish bilinguals was constant across Spanish and English experimental contexts. Likewise, for the domain of objects, Athanasopoulos (2007) found no difference in object categorization preferences for Japanese-English bilinguals performing the task in different linguistic contexts.

The question of whether bilinguals align with monolingual speakers is related to, but yet distinct from, the question of cognitive flexibility, as a bilingual individual may exhibit distinct patterns of, for instance, categorisation, in each of their languages, without necessarily aligning with monolingual speakers of either language. Research comparing bilingual and monolingual cognitive patterns often shows that the extent to which they align is dependent on a number of individual background variables. One important such variable is language proficiency (L1 and/or L2), the assumption being that more proficient individuals are more likely to behave like native monolingual speakers than less proficient individuals. Indeed, in bilinguals who are still developing competence in the L2, increasing proficiency has been shown to underpin cognitive restructuring or shift towards the L2 (e.g., Athanasopoulos, 2007; Athanasopoulos & Kasai, 2008; Kurinski & Sera, 2011; Park & Ziegler, 2014).

Language proficiency has also been shown to be a significant factor in predicting bilingual speakers' temporal perspective. Based on the observation that, contrary to English, time-moving metaphors in Mandarin ('the deadline is approaching') are more frequent than ego-moving metaphors ('we are approaching the deadline'), Lai and Boroditsky (2013) asked participants questions where the answer required adopting either a time-moving or an ego-moving perspective. Their results showed that English speakers were indeed more likely to adopt the Ego-moving perspective than were Mandarin speakers, and furthermore that bilinguals' temporal perspective was predicted by language proficiency, such that those who were more proficient in Mandarin were less likely to take an ego-moving perspective, and those who were more proficient in English were more likely to adopt such a perspective.

It should be noted however that many studies do not find effects of language proficiency on cognitive restructuring (see e.g. Athanasopoulos, Dering, Wiggett, Kuipers & Thierry, 2010; Bylund & Athanasopoulos, 2014b; Cook, Bassetti, Kasai, Sasaki, & Takahashi, 2006 who all found effects of length of residence in the L2 speaking country instead of L2 proficiency). As suggested by Bylund and Athanasopoulos (2014a), these mixed findings may in part be ascribed to the different test formats used for assessing proficiency. One other reason may be that the bilingual samples tested do not show requisite variability in their L2 proficiency to render any correlation, regression, or factorial analysis significant. For instance, Lai and Boroditsky (2013) investigated the impact of horizontal and vertical metaphors on temporal sequencing in Mandarin-English bilinguals. Consistent with Mandarin, which adopts a vertical system for sequencing events, Mandarin-English bilinguals placed future events below and past events above more frequently when tested in Taiwan than in the USA, demonstrating a robust effect of experimental context. Language proficiency, on the other hand, did not predict behaviour in this task, possibly because all participants were very advanced in both of their languages.

Another important variable is age of L2 acquisition, which has been found to influence cognitive restructuring in at least two different ways. First, it may exert an indirect effect via L2 or L1 proficiency: age of L2 acquisition is often negatively correlated with L2 proficiency, and positively correlated with L1 proficiency. Proficiency, as shaped by age of acquisition, may then consequently restructure cognitive patterns, as revealed when effects of proficiency are controlled for (note that in these instances, age effects are expected to be found for naturalistic learners rather than for instructed learners). Second, age of acquisition may modulate the extent to which language-specific conceptual representations are merged or kept separate. For example, Kersten et al. (2010) found that early bilinguals were more prone to exhibit similar event categorisation patterns when operating in either of their languages, whereas late bilinguals would categorise events differently depending on the language context.

In the domain of time, age of L2 acquisition is the only variable that has been found to correlate significantly with degree of cognitive restructuring. Specifically, Boroditsky (2001) found that Mandarin speakers who started learning English earlier in life exhibited less of a vertical bias in a semantic verification task with priming (recall that vertical spatiotemporal metaphors proliferate in Chinese but not English). Results showed a positive correlation between vertical bias and age of L2 acquisition, such that the greater the vertical bias the later the age of L2 acquisition of the participant. In other words, the longer a participant was speaking only Mandarin in their language development, the greater their vertical bias, showing an instance of L2 effect on a temporal dimension -verticality- exclusively encoded in the L1. Subsequent studies challenged Boroditsky's (2001) main claims of a verticality bias in Mandarin time conceptualisation (e.g. Chen, 2007; January & Kako, 2007; Tse & Altarriba, 2008) but did not specifically focus on bilinguals/L2 users per se in their investigations.

Finally, frequency of use of the relevant languages appears to also exert an influence on bilingual cognitive restructuring, with several studies showing that the more often a bilingual speaker uses a given language the more likely they are to converge with monolingual speakers of that language on a cognitive task (Athanasopoulos, Damjanovic, Krajciova & Sasaki, 2011; Bylund & Athanasopoulos, 2014b, 2015; Bylund, Athanasopoulos, & Oostendorp, 2013; Park & Ziegler, 2014). To our knowledge, no study has examined the role of this variable in bilingual temporal cognition.

Aims and scope of the present study

The current study sets out to throw light on how the acquisition of a new language influences time perception. While some progress has been made on understanding the bilingual processing of temporal sequences (e.g., Boroditsky, 2001; Miles et al., 2011), little is still known about bilingual perception of temporal spans (Bylund & Athanasopoulos, 2017), and the factors that modulate temporal cognition in general in bilinguals. To address this, we examined duration reproduction in Spanish-Swedish functional bilinguals, with the intention to study, first, the extent to which L2 time perception patterns converge with those of native speakers, and second, the factors that underlie such (non)convergence. In order to obtain a nuanced picture of bilingual time perception patterns, we tested duration reproduction for both line and container animations in a between-subjects design, thus permitting an assessment of time perception of not only spatial depictions specific to L2 temporal metaphors (i.e., lines), but also depictions of L1-specific temporal metaphors (i.e., containers), in an all Swedish (L2) testing language context. This allows us to examine both internalisation of novel time concepts prevalent in the L2 (duration as distance, lines experiment), as well as any L2 effects on sensitivity to time concepts prevalent in the L1 (duration as quantity, containers experiment).

We can make at least three different predictions regarding our bilinguals, interpreted differently for the two experiments, lines (the L2 pattern to be learned), and containers (the pre-existing L1 bias). First, bilinguals' time estimation patterns may resemble those of monolingual speakers of their L1. For lines, this would suggest that bilinguals have not internalized at all the L2 pattern of thinking about time as distance. For containers, this would suggest that there is no traceable conceptual attrition – patterns established by the L1 remain unchanged and prevalent. These outcomes would be compatible with the entrenchment hypothesis that cognitive patterns, once established by the L1, are exceptionally resistant to restructuring by an L2 (cf. Slobin, 1996). A second possible prediction is that bilinguals' time estimation patterns may resemble those of monolingual speakers of their L2. In the case of lines, this would mean full internalization of the L2 way of thinking about duration as distance travelled. In the case of containers, it would mean that the L1 pattern of duration estimation is no longer prevalent, or is less salient, in bilingual cognition, constituting evidence of L1 conceptual weakening. These outcomes would favour the restructuring hypothesis. A third possibility is convergence, a compromise between resistance and restructuring, indexed by an 'in-between' pattern that resembles neither the L1 nor the L2. For lines, this would mean partial internalisation of the L2 pattern, but not to a native-like extent. For containers, it would mean partial attrition of the L1 pattern.

Because of our design that tests L2 internalization and L1 attrition separately, we may find that more than one of the above predictions is supported, to different degrees for each of the experiments. A necessary additional analytical step then, is to explore what accounts for the variance displayed in bilingual cognitive behaviour, by assessing the relative influence of each of the biographical variables of interest, specifically age of L2 acquisition, L2 proficiency, frequency of use, and length of residence in the L2 speaking country, on bilingual individuals' temporal cognition patterns.

Method

Participants

Eighty L1 Spanish – L2 Swedish bilinguals living in Sweden were evenly allocated to either the Lines or the Containers experiment. Five participants were excluded from the analysis due to poor performance in the experiments, presumably due to impatience and/or fatigue, typical of psychophysical tasks of this type. Specifically, following Casasanto and Boroditsky (2008) and Casasanto (2005), participants were removed if they estimated distance instead of time, or their overall duration estimations were markedly inaccurate (if the slope of the correlation between actual and estimated duration was <.5). The remaining participants had the following characteristics:

For the lines experiment (n = 39), the bilinguals' length of residence in Sweden was 24.7 years on average (SD 7.4, range 10-45 years). Their average age of acquisition of L2 Swedish was 9.6 years (SD 7.8, range 1-28 years). Their average Swedish proficiency score as measured by a Cloze test (see Materials) was 20.28/42 (SD 5.65, range 7-30).[1] They used Swedish for communicative purposes 76.3% (SD 14.3%) of the time, and Spanish for the remainder.

For the containers experiment (n = 36), the bilinguals had lived in Sweden for an average of 21.9 years (SD 6.7, range 8-39 years). On average, they had started learning Swedish at 11 years (SD 9.1, range 1-31 years). Their average Swedish proficiency score as measured by a Cloze test (see Materials) was 18.97/42 (SD 5.9, range 6-29). In their everyday lives, they used Swedish 73.1% (SD 18.9%) of the time, and Spanish for the remainder.

The functionally monolingual native speakers of Swedish (n = 35) and Spanish (n = 35) analysed in Bylund and Athanasopoulos (2017) were used as control groups for comparison purposes here (Lines experiment: Swedish = 17; Spanish = 18. Containers experiment: Swedish = 18; Spanish = 17).

Materials

Lines Experiment

Casasanto's (2005) growing lines experiment was used to elicit duration reproduction for spatial depictions analogous to L2 temporal metaphors (i.e., Swedish distance-based). In this experiment, 9 different line distances, ranging from 100 to 500 pixels (with 50 pixels increments), and 9 different line durations, ranging from 1000 milliseconds (ms) to 5000 ms (500 ms increments) were fully crossed. This resulted in a total of 81 unique line stimuli. Lines were black and grew from left to right against a white background.

Containers Experiment

Casasanto's (2005) filling containers experiment was used to elicit duration reproduction for spatial depictions analogous to L1 temporal metaphors (i.e., Spanish quantity/volume-based). In this experiment, 9 different fill levels, ranging from 100 to 500 pixels (with 50 pixels increments), and 9 different container durations, ranging from 1000 ms to 5000 ms (500 ms increments) were fully crossed, yielding a total of 81 unique filling container stimuli. Containers were comprised of a 600 x 600 pixels hollow frame, and were filled in black from the bottom upwards against a white background.

A language background questionnaire was also administered in both experiments, in order to collect information about acquisition history and usage of the bilingual participants, and to ensure that the monolingual controls were not using other languages than their L1. Bilingual participants in both experiments also took a Swedish proficiency test. A cloze test was chosen for this purpose, since this format yields reliable measures of language proficiency across a range of different learner levels (McNamara, 2000; Tremblay, 2011). The test consisted of an around 300-word long text, in which every 7th word had been removed (Platzack, 1974).

Procedure and Design in both experiments

Participants were tested individually on a 15.6” laptop. A Swedish native speaker tested the bilinguals and the Swedish monolinguals, and a Spanish native speaker tested the Spanish monolinguals. The procedure and design were identical to Casasanto and associates (Bylund & Athanasopoulos, 2017; Casasanto, 2005; 2008; Casasanto et al., 2004;) as follows:

Each stimulus (growing line or filling container) was presented twice, resulting in a total of 162 trials per experiment. Half of the times the participants estimated duration, and the other half they estimated displacement (distracter task). Before each stimulus, a verbal prompt indicated whether duration or displacement was to be estimated (the word ‘duration’ or ‘displacement’ in Swedish for the bilinguals and the Swedish monolinguals; the same words in Spanish for the Spanish monolinguals). Participants reproduced duration by clicking the computer mouse once to mark the onset, waited the appropriate time, and then clicked again to mark the terminus. Presentation orders were fully randomised.

In accordance with previous research, we focused on the so-called medium, hard-to-process stimuli in our analysis, which are the five intermediate durations (200 ms to 400 ms) and physical lengths/volume levels (200 pixels to 400 pixels), since it is for these that language-specific spatial interference is found (Bylund & Athanasopoulos, 2017; Casasanto, 2005; Casasanto et al., 2005). This yielded a total of 25 trials analysed per participant. Linear mixed-effects models were run with estimated duration as the dependent variable. For the group comparisons, stimulus displacement and group were included as predictor variables, and actual stimulus duration as a co-variate. Random intercepts were included for participants and items, as justified by the maximal model that converged (in the case of the Lines experiment, the model that converged included only participants in the random effect structure). We used treatment coding, re-refencing the model to each group for contrasts of interest.

For the analyses of bilingual background variables, we tested for the inclusion of predictors to see which would improve the model fit significantly. In these comparisons, the baseline model included actual stimulus duration as a covariate, and the interaction between stimulus displacement and age of L2 acquisition as predictor, based on previous findings that showed age of L2 acquisition to be the only variable to correlate significantly with degree of restructuring in bilingual temporal cognition (Boroditsky, 2001). We then added the interaction between displacement and each new predictor (L2 proficiency, frequency of L2 use, length of residence in the L2 speaking country) and performed model comparisons to discover the model that would best fit the data. The maximal model that converged included random intercepts for participants and items in the Containers Experiment, and only participants in the Lines Experiment. All analyses were run in R (version 4.1.0, The R Foundation for Statistical Computing, 2021).

Results

Lines experiment

Group comparisons: Results indicated that, unsurprisingly, the covariate of actual duration was significantly related to duration estimation across all groups ($\beta = 0.77$, $SE = 0.02$, $t = 46.31$, $p < 0.01$; values were identical across the three models with each group as reference). Figure 1 shows that as predicted, Swedish monolinguals were affected to a significantly greater extent by line displacement compared to their monolingual Spanish peers ($\beta = 1.34$, SE = 0.49, t = 2.76, $p < 0.01$, model with Spanish monolinguals as reference; values were identical in the model with Swedish monolinguals as reference except β = -1.34 and t = -2.76). Bilinguals did not differ significantly from Swedish monolinguals (t = 1.33) and differed marginally from Spanish monolinguals (t = -1.92) (values were identical in the models with Swedish and Spanish monolinguals as reference, except t = -1.33 and t = 1.92 respectively).

<Insert Figure 1 about here>

Bilingual Background Variables: The baseline model with only the interaction between displacement and age of L2 acquisition was not statistically significant. Adding the interaction between displacement and L2 proficiency yielded a statistically significant model, such that the higher the proficiency in their L2 Swedish, the more bilinguals were affected by line displacement (β = 0.05, SE = 0.02, t = 2.82, $p < 0.01$), while age of L2 acquisition did not interact significantly with displacement. Critically, this model was also a statistically better fit for the data than the baseline model, $x^2 = 7.76$, $p < 0.01$. Adding interaction terms between displacement and each of the other two variables (frequency of L2 use and length of residence in the L2 country) did not significantly improve the model fit (*p*s > .10).

Containers experiment

Group comparisons: As in the lines experiment, the covariate of actual duration exerted a significant effect on duration estimation, as expected (β = 0.75, SE = 0.02, t = 48.21, $p<0.01$; values were identical across the three models with each group as reference). Figure 2 shows that according to our predictions, Spanish monolinguals were affected to a significantly greater extent by container displacement compared to their Swedish monolingual peers (β = -

1.79, SE = 0.44, t = -4.04, p < 0.01, model with Spanish monolinguals as reference; values were identical in the model with Swedish monolinguals as reference, except β = 1.79 and t = 4.04). Bilinguals differed significantly from both Swedish monolinguals (β = -0.89, SE = 0.38, t = -2.34, p < 0.02) and Spanish monolinguals (β = 0.90, SE = 0.38, t = 2.37, p < 0.02) (values were identical in the models with Swedish and Spanish monolinguals as reference, except β = 0.89, t = 2.34, and β = -0.90, t = -2.37, respectively).

<Insert Figure 2 about here>

Bilingual background variables: The baseline model with the interaction between displacement and age of L2 acquisition was statistically significant, such that bilinguals that learned the L2 Swedish later in life were likely to exhibit greater spatial interference in their duration estimation of filling containers (β = 0.05, SE = 0.02, t = 3.60, p < 0.01). Adding interaction terms between displacement and each of the remaining biographical variables (L2 proficiency, frequency of L2 use, length of residence in the L2 speaking country) did not significantly improve the model fit (*p*s > .10).

Discussion

The current paper, by employing a psychophysical time estimation task, set out to test the degree of internalisation of a L2 pattern to conceptualise duration, the degree of resistance to restructuring of L1 patterns of duration conceptualisation, and the factors that may account for bilingual cognitive processing in the domain of time. Because of our design that examined L2 internalization and L1 retention in two separate experiments, it was possible to shed light on these phenomena independently. Our group analyses found evidence of L2 internalisation as well as L1 conceptual weakening. As can be seen in the statistical analysis, and in figures 1 and 2, the bilingual groups displayed an ‘in-between’ pattern in the degree to which their time estimations were affected by spatial displacements, relative to the monolingual datapoints. Unlike previous studies that focused on categorisation, where such mid-points could either reflect chance behaviour on a binary choice, or a bimodal response pattern between two categorisation options, here the dependent variable was continuous, representing variation across the bilingual sample in each of the two spatial interference conditions.

The question of interest then was what feature of the bilingual experience may account best for the variation displayed. Our findings revealed that L2 internalisation was modulated by L2 proficiency, such that the more proficient the bilingual individual became in the L2, the more they were affected by the spatial dimension prevalent in L2 spatiotemporal metaphoric structures (in this case, time as linear distance). This finding is in line with a plethora of previous findings showing effects of increasing L2 proficiency on changing cognitive patterns in the individual in various perceptual domains (Bylund & Athanasopoulos, 2014a).

L1 conceptual attrition on the other hand appears to be a function of the age of L2 acquisition of the bilinguals that took part in our study. The later the individual began acquiring their L2, the more L1-like their time estimation was of stimuli representing the dominant metaphoric structure of their L1 (in this case, time as quantity). Our finding that age of L2 acquisition predicts weakening of the L1 conceptual pattern is fully in line with previous literature reporting similar findings from the time domain. Boroditsky (2001) reported that in Mandarin-English bilinguals, the vertical bias towards temporal sequencing encoded in the L1 was greater for bilingual speakers who started learning L2 English later in life. The language of instruction in that experiment was the L2, similarly to our experiment. Therefore, in both studies, this is an instance of an L2 effect on a temporal dimension - verticality, in the case of Mandarin-English bilinguals; duration-as-quantity, in the case of Spanish-English bilinguals - preferentially encoded in the L1. Parallel to Boroditsky’s (2001) findings, the propensity to think about duration as quantity was related to the length of monolingual Spanish experience (before any Swedish was learned).

What accounts then, for the effects of L2 proficiency and age of L2 acquisition on the psychophysical experience of time in our bilingual sample? Many recent studies have attempted to contextualise their findings in the theoretical framework of associative/attentional learning (Athanasopoulos, Damjanovic, Burnand & Bylund, 2015b; Casasanto, 2008; Kersten et al., 2010; Lai et al., 2014) and have even posited such a framework as an ideal candidate for delineating the mechanism underpinning bilingual cognition effects involving language-thought interactions (Bylund & Athanasopoulos, 2014a). Our study presents strong empirical evidence to support the postulation of such a mechanism in bilingual cognition. In line with this literature (see e.g. Kersten et al., 2010), the age effect apparent here could be interpreted under an associative learning account. Early bilinguals statistically receive a mixture of language inputs highlighting both duration-as-distance and duration-as-quantity time schemata, resulting in more equally weighted cues for duration estimation. Later acquisition of the L2, on the other hand, means that individuals start out by receiving input that is statistically biased towards the duration-as-quantity metaphor, with later acquisition introducing more frequent cues towards duration-as-distance, allowing more opportunity for the L1 based metaphoric pattern to be established. However, it remains to be seen whether the nature of this relationship is linear, or whether there is a threshold or breakpoint that could be indicative of a critical or sensitive period for cognitive restructuring in the bilingual individual. Future studies may employ a systematic breakpoint identification procedure with a larger sample of participants in this age-range along the lines demonstrated in Vanhove (2013) and Veríssimo, Heyer, Jacob and Clahsen (2018).

The findings from the lines experiment corroborate the assumption, under an associative learning account, that increasing expertise in a particular language is likely to be a key aspect of consolidation of language-specific spatiotemporal representations in memory. Here, the degree of L2 internalisation was reliably predicted by the individual's L2 proficiency score. As L2 proficiency increases, links between physical length and distance time metaphors appear to also be strengthened.

Furthermore, the lack of an observable effect of length of stay in the L2 speaking country is not surprising, given the profile of our participants with regards to this variable. Specifically, studies document effects of length of stay on bilingual cognition in participant samples residing in the L2 speaking country typically for periods between 5 months to 8 years (e.g. Athanasopoulos et al., 2010; Cook et al., 2006). Boroditsky (2001), in contrast, found no significant effects of length of stay on temporal cognition in her sample of Mandarin-English bilinguals who had at least 10 years of residence in an English-speaking setting. This seemingly contrasting set of findings from the afore-mentioned studies is in fact compatible with the fact that a significant correlation between L2 ultimate attainment/L1 attrition and length of residence in the L2 speaking country is primarily observed during the first decade in the L2 context, and past this time frame length of residence exerts zero or little effect (e.g., Bylund, Abrahamsson & Hyltenstam, 2012; DeKeyser, 2000; Schmid, 2011). Against this background, and given the longer-than-10-years average length of stay of our bilingual samples in both experiments, it is not surprising that, like Boroditsky (2001), we do not find a significant effect of this variable. Further systematic research is necessary in order to establish whether effects of length of residence may level out in cases of prolonged exposure in the field of bilingual cognition.

Finally, our findings are fully compatible with a theoretical framework of cognition that assumes a distributed representational system that places the interaction between internal mental computations and their embedding in external environmental features at its core[2]. Specifically, theories of distributed cognition assume that cognitive activity is distributed beyond the “boundaries of skin and skull” (Clark & Chalmers, 1998:7). That is, cognitive processing is not confined to the mental computations occurring exclusively in the mind/brain, but it is rather a two-way interaction between the individual or a group of individuals, and the external environment they operate in (Clark, 2017). All the components, internal and external, of this extended cognitive system are actively engaged in cognitive processing. Instantiations of this framework have been traditionally documented in the way groups of individuals cooperate for the purposes of spatial navigation (Hutchins, 1994), or in the way pilots interact with the instruments in their cockpit to calculate airspeed (Hutchins, 1995).

Our findings here, along with previous findings from the same experimental paradigm (Casasanto, 2005; Bylund & Athanasopoulos, 2017), show that space, as an external stimulus, plays a crucial role in the way individuals construct their mental representations of time, on the go, as it were. As such, our study aligns with the framework of distributed cognition, and furthermore shows that such interactions between external stimulus and internal mental computation are highly malleable, as revealed in the changing cognitive patterns of bilingual individuals. This is also anticipated by the framework of distributed cognition (e.g. Hollan, Hutchins & Kirsh, 2000), which assumes that external stimuli are inextricably integrated with internal computations, such that if the internal computation remains the same (in this case, estimating duration) but the external features change (in this case, growing lines versus filling containers), behaviour also may change completely. In this light, the extant findings highlight the role of bilingualism on the distributed nature of cognition, by showing that experience with a second language appears to be an important determinant of this malleability in the system.

To conclude, our study examined a particular domain of experience that is typically encoded differently in the two languages of the bilingual individual. Our results suggest that learners not only acquire new links between language-specific metaphors and the psychophysical experience of duration, but may also unlearn or partially develop the original ones, as a function of the age at which they started learning their L2. Looking at time perception in two different spatial dimensions, each exclusively related to one frequent type of spatiotemporal metaphor in each of the bilingual's languages, yields a unique affordance to investigate bilingual cognition of both L1-salient and L2-salient conceptual representations. We can then become more confident that in isolating those variables that modulate cognitive patterns, we are closer to uncovering the critical factors that govern both internalisation of new concepts, as well as the fate of the existing concepts, in development. Our findings have thus opened up the way for further systematic investigation of these phenomena by ferreting out further the nature of the influence of critical factors of L2 development and cognitive restructuring.

Competing interests: The authors declare none

References

Abrahamsson, N and Hyltenstam, K (2009) Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning* 59, 249–306. https://doi.org/10.1111/j.1467-9922.2009.00507.x

Athanasopoulos, P (2007) Interaction between grammatical categories and cognition in bilinguals: The role of proficiency, cultural immersion, and language of instruction. *Language and Cognitive Processes* 22, 689–699. https://doi.org/10.1080/01690960601049347

Athanasopoulos, P and Kasai, C (2008) Language and thought in bilinguals: The case of grammatical number and nonverbal classification preferences. *Applied Psycholinguistics* 29, 105–123. https://doi.org/10.1017/S0142716408080053

Athanasopoulos, P, Dering, B, Wiggett, A, Kuipers, J and Thierry, G (2010) Perceptual shift in bilingualism: Brain potentials reveal plasticity in pre-attentive colour perception. *Cognition* 116, 437–443.

Athanasopoulos, P (2011) Cognitive restructuring in bilingualism. In Pavlenko, A (ed), *Thinking and speaking in two languages*. Bristol, UK: Multilingual Matters, pp. 29–65.

Athanasopoulos, P, Damjanovic, L, Krajciova, A and Sasaki, M (2011) Representation of colour concepts in bilingual cognition: The case of Japanese blues. *Bilingualism: Language and Cognition* 14, 9–17.

Athanasopoulos, P, Bylund, E, Montero-Melis, G, Damjanovic, L, Schartner, A, Kibbe, A, Riches, N and Thierry, G (2015a) Two languages, two minds: Flexible cognitive processing driven by language of operation. *Psychological Science* 26, 518–526. https://doi.org/10.1177/0956797614567509

Athanasopoulos, P, Damjanovic, L, Burnand, J and Bylund, E (2015b) Learning to think in a second language: Effects of proficiency and length of exposure in English learners of German. *Modern Language Journal* 99, 138-153. https://doi.org/10.1017/S1366728909990046

Athanasopoulos, P, Bylund, E and Casasanto, D (2016) Special issue: New and interdisciplinary approaches to linguistic relativity. *Language Learning* 66, 471-730.

Boroditsky, L (2001) Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology* 43, 1–22. https://doi.org/10.1006/cogp.2001.0748

Boroditsky, L (2011) How languages construct time. In Dehaene, S and Brannon, E (eds), *Space, Time and Number in the Brain*. Amsterdam: Elsevier, pp. 333-341.

Bylund, E, Abrahamsson, N and Hyltenstam, K (2012) Does L1 maintenance hamper L2 nativelikeness? A study of L2 ultimate attainment in early bilinguals. *Studies in Second Language Acquisition* 34, 215–241.

Bylund, E, Athanasopoulos, P and Oostendorp, M (2013) Motion event cognition and grammatical aspect: Evidence from Afrikaans. *Linguistics* 51, 929–955.

Bylund, E and Athanasopoulos, P (2014a) Linguistic Relativity in SLA: Toward a New Research Program. *Language Learning* 64, 952–985. https://doi.org/10.1111/lang.12080

Bylund, E and Athanasopoulos, P (2014b) Language and thought in a multilingual context: The case of isiXhosa. *Bilingualism: Language and Cognition* 17, 431–443.

Bylund, E and Athanasopoulos, P (2015) Motion event categorisation in a nativised variety of South African English. *International Journal of Bilingual Education and Bilingualism* 18, 588–601. https://doi.org/10.1080/13670050.2015.1027145

Bylund, E and Athanasopoulos, P (2017) The Whorfian time warp: Representing duration through the language hourglass. *Journal of Experimental Psychology: General* 146, 911–916. https://doi.org/10.1037/xge0000314

Bylund, E, Hyltenstam, K and Abrahamsson, N (2021) Age of acquisition – not bilingualism – is the primary determinant of less than nativelike L2 ultimate attainment. *Bilingualism: Language and Cognition* 24, 18–30. https://doi.org/10.1017/S1366728920000188

Casasanto, D, Boroditsky, L, Phillips, W, Greene, J, Goswami, S, Bocanegra-Thiel, S, Santiago-Diaz, I, Fotakopolou, O, Pita, R and Gil, D (2004) How deep are effects of language on thought? Time estimation in Speakers of English, Indonesian, Greek and Spanish. In Forbus, K, Gentner, D and Regier, T (eds), *Proceedings of the 26th Annual Conference Cognitive Science Society*. Lawrence Erlbaum, pp. 575–580. http://pubman.mpdl.mpg.de/pubman/faces/viewItemOverviewPage.jsp?itemId=escidoc:1555664

Casasanto, DJ (2005) *Perceptual foundations of abstract thought*. Massachusetts Institute of Technology. http://dspace.mit.edu/handle/1721.1/34129

Casasanto, D (2008) Who’s afraid of the big bad Whorf? Crosslinguistic differences in temporal language and thought. *Language Learning* 58, 63–79.

Casasanto, D and Boroditsky, L (2008) Time in the mind: Using space to think about time. *Cognition* 106, 579–593. http://doi.org/10.1016/j.cognition.2007.03.004

Chen, JY (2007) Do Chinese and English speakers think about time differently? Failure of replicating Boroditsky (2001). *Cognition* 104, 427–436.

Clark, A and Chalmers, D (1998) The Extended Mind. *Analysis* 58(1), 7–19. http://www.jstor.org/stable/3328150

Clark, A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* 36, 181–204. https://doi.org/10.1017/S0140525X12000477

Clark, A (2017) Embodied, Situated, and Distributed Cognition. In Bechtel, W and Graham, G (eds), *A Companion to Cognitive Science*, Blackwell Publishing Ltd., pp. 506-517.

Clark, HH (1973) Space, time, semantics, and the child. In Moore, T (ed), *Cognitive development and the acquisition of language*. Academic Press, pp. 27–63.

Cook, VJ (2002) Background to the L2 user. In Cook, VJ (ed), *Portraits of the L2 User*. Clevedon: Multilingual Matters, pp. 1–28.

Cook, V, Bassetti, B, Kasai, C, Sasaki, M and Takahashi, JA (2006) Do bilinguals have different concepts? The case of shape and material in Japanese L2 users of English. *International Journal of Bilingualism* 10, 137–152. https://doi.org/10.1177/13670069060100020201

DeKeyser, R (2000) The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition* 22, 499–533.

Filipović, L (2011) Speaking and remembering in one or two languages: Bilingual vs. monolingual lexicalization and memory for motion events. *International Journal of Bilingualism* 15, 466–485. https://doi.org/10.1177/1367006911403062

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

Hohwy, J (2013) *The Predictive Mind*. Oxford: Oxford University Press.

Hollan, J, Hutchins, E and Kirsh, D (2000) Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction* 7, 174–196.

Hutchins, E (1994) *Cognition in the Wild*. MIT Press, Cambridge, MA.

Hutchins, EL (1995) How a cockpit remembers its speed. *Cognitive Science* 19, 265–288.

January, D and Kako, E (2007) Re-evaluating evidence for linguistic relativity: Reply to Boroditsky (2001). *Cognition* 104, 417–426.

Kanai, R, Komura, Y, Shipp, S and Friston, K (2015) Cerebral hierarchies: Predictive processing, precision and the pulvinar. *Philosophical Transactions of the Royal Society B* 370, 20140169. https://doi.org/10.1098/rstb.2014.0169

Kersten, AW, Meissner, CA, Lechuga, J, Schwartz, BL, Albrechtsen, JS and 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, 638–653. https://doi.org/10.1037/a0020507

Kurinski, E and Sera, MD (2011) Does learning Spanish grammatical gender change English-speaking adults' categorization of inanimate objects? *Bilingualism: Language and Cognition* 14, 203–220. https://doi.org/10.1017/S1366728910000179

Lai, VT and Boroditsky, L (2013) The immediate and chronic influence of spatio-temporal metaphors on the mental representations of time in English, Mandarin, and Mandarin-English speakers. *Frontiers in Psychology* 4, 142. https://doi.org/10.3389/fpsyg.2013.00142

Lai, V, Rodriguez, G and Narasimhan, B (2014) Thinking-for-speaking in early and late bilinguals. *Bilingualism: Language and Cognition* 17, 139-152. doi:10.1017/S1366728913000151

Lakoff, G and Johnson, M (1980) *Metaphors we live by*. University of Chicago Press.

Lucy, JA (2004) Language, culture, and mind in comparative perspective. In Achard, M and Kemmer, S (eds), *Language, Culture, and Mind*. Stanford: Center for the Study of Language and Information Publications [distributed by the University of Chicago Press], pp. 1–21.

Lucy, JA (2016) Recent advances in the study of linguistic relativity in historical context: A critical assessment. *Language Learning* 66, 487-515.

Lupyan, G (2012) Linguistically modulated perception and cognition: The Label-Feedback Hypothesis. *Frontiers in Psychology* 3. https://doi.org/10.3389/fpsyg.2012.00054

Lupyan, G and Clark, A (2015) Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science* 24, 279–284. https://doi.org/10.1177/0963721415570732

McNamara, T (2000) *Language testing*. Oxford University Press.

Miles, LK, Tan, L, Noble, GD, Lumsden, J and Macrae, CN (2011) Can a mind have two time lines? Exploring space-time mapping in Mandarin and English speakers. *Psychonomic Bulletin & Review* 18, 598–604. https://doi.org/10.3758/s13423-011-0068-y

Núñez, R and Cooperrider, K (2013) The tangle of space and time in human cognition. *Trends in Cognitive Sciences* 17, 220–229. https://doi.org/10.1016/j.tics.2013.03.008

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

Park, HI and Ziegler, N (2014) Cognitive shift in the bilingual mind: Spatial concepts in Korean–English bilinguals. *Bilingualism: Language and Cognition* 17, 410–430. https://doi.org/10.1017/S1366728913000400

Pavlenko, A (1999) New approaches to concepts in bilingual memory. *Bilingualism: Language and Cognition* 2, 209–230.

Platzack, C (1973) *Språket och läsbarheten*. Lund: Gleerup.

Schmid, M (2011) *Language attrition*. Cambridge, UK: Cambridge University Press.

Slobin, DI (1996) From "thought and language" to "thinking for speaking". In Gumperz, JJ and Levinson, SC (eds), *Rethinking linguistic relativity*. Cambridge: Cambridge University Press, pp. 70–96.

Traugott, EC (1978) On the expression of spatio-temporal relations in language. *Universals of Human Language* 3, 369–400.

Tremblay, A (2011) Proficiency assessment standards in second language acquisition research: "Clozing" the Gap. *Studies in Second Language Acquisition* 33, 339-372. doi:10.1017/S0272263111000015

Tse, CS and Altarriba, J (2008) Evidence against linguistic relativity in Chinese and English: A case study of spatial and temporal metaphors. *Journal of Cognition and Culture* 8, 335–357.

Vanhove, J (2013) The Critical Period Hypothesis in Second Language Acquisition: A Statistical Critique and a Reanalysis. PLoS ONE 8(7): e69172. doi:10.1371/journal.pone.0069172

Veríssimo, J, Heyer, V, Jacob, G and Clahsen, H (2018) Selective Effects of Age of

Acquisition on Morphological Priming: Evidence for a Sensitive Period. *Language Acquisition* 25, 315-326. DOI: 10.1080/10489223.2017.1346104

von Humboldt, W (1963) *Humanist without portfolio: An anthology of the writings of Wilhelm von Humboldt*. Detroit, MI: Wayne State University Press.

Whorf, BL (1956) *Language, thought, and reality: Selected writings (edited by J. B. Carroll)*. Cambridge: MIT Press.

Winawer J, Witthoft N, Frank MC, Wu L, Wade AR and Boroditsky L (2007) Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the USA* 104(19):7780-5. doi: 10.1073/pnas.0701644104.

Yang, W and Sun, Y (2016) A monolingual mind can have two time lines: Exploring space-time mappings in Mandarin monolinguals. *Psychonomic Bulletin & Review* 23, 857–864.

https://doi.org/10.3758/s13423-015-0964-

Footnotes

[1] Previous studies using the same cloze test have found that Swedish L1 speakers on average score around 30 to 35 (Abrahamsson and Hyltenstam, 2009). L2 speakers of Swedish with early ages of acquisition (3-8 years) obtain around 25 (Bylund, Abrahamsson and Hyltenstam, 2021). An average score of around 20 can thus be expected from a mixed group of early and late learners.

[2] We are grateful to Reviewer 2 for bringing our attention to this theoretical framework and for suggesting the compatibility of our findings with its basic tenets.

Figure 1. The interaction between estimated duration (in milliseconds, vertical axis) and spatial displacement (in pixels, horizontal axis) as a function of Group in the Lines experiment.

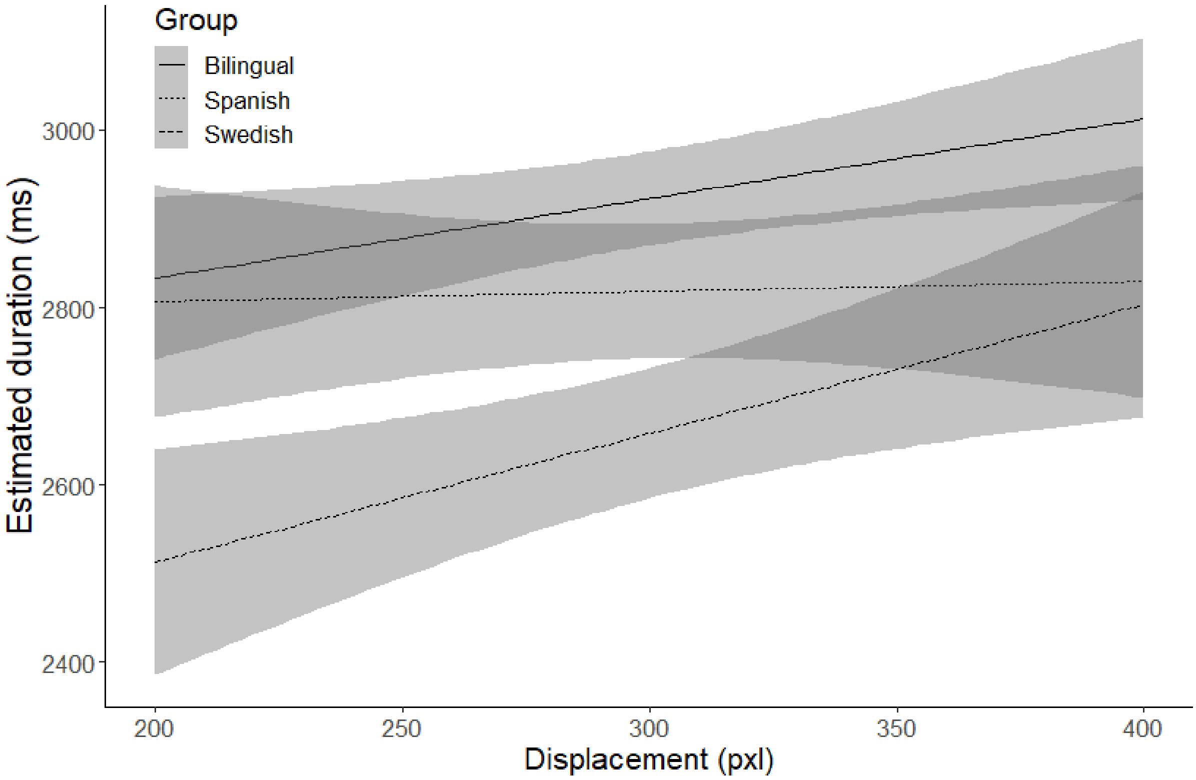

Figure 2. The interaction between estimated duration (in milliseconds, vertical axis) and spatial displacement (in pixels, horizontal axis) as a function of Group in the Containers experiment.