Proceeding of ICIHCS 2025 Symposium: The Dialogue Between Tradition and Innovation in Language Learning
DOI: 10.54254/2753-7064/2025.HT30156

Disappearing or Hidden? Reconciling VIP and ERP Findings on Linguistic Relativity

Zhuang Yang

Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom
s2658083@ed.ac.uk

Abstract. The linguistic relativity hypothesis (LR) believes that different language speakers may conceptualize the world differently. In recent decades, the debate has shifted from whether language shapes thought to when and how such modulation occurs, especially in temporal domains. The review examines how language influences thought and behavior by comparing findings from Verbal Interference Paradigms (VIP) and Event-Related Potential (ERP) studies. Some LR effects have been posited to occur via online use of language. Evidence for this comes from VIP, in which language resources are occupied by a concurrent task. This evidence has been interpreted to indicate that without access to language at the moment, speakers of different languages think the same. However, ERP evidence shows that neural differences often persist even when overt behavior shows no effect. This new evidence suggests that apparent null results may reflect methodological thresholds rather than the absence of an influence of language on thought. Together, these findings suggest the importance of distinguishing cognitive processing from behavioral outcomes when evaluating the research scope of LR.

Keywords: Linguistic relativity, Verbal Interference Paradigms, Event-related potentials, Cognitive linguistics, Language and thought

1. Introduction

The relationship between language and cognition has long been an attractive but contentious topic in the contemporary academic field. One of the most fiercely debated topics must be the Sapir–Whorf hypothesis (Linguistic Relativity, LR). It begins with Whorf's metaphor that the world is perceived as a "kaleidoscopic flux of impressions," which implies that our cognitive framework is dependent upon linguistic knowledge [1]. With the development of experimental psychology and nativists, scholars reject such strong determinism interpretation but argue a more moderate stance. Rather than confining cognition, language is proposed to modulate cognition by directing attention and biasing perception. Such relativism has been supported empirically, especially in domains of time [2, 3], number [4, 5], motion events [6, 7], and grammatical gender [8, 9]. Yet critics often point to null results showing universal tendencies or contextual confounds [10, 11]. The mixed evidence raises two key questions: (a) Does language modulate cognition, or are the observed effects experimental artifacts? (b) If it does, when and how do these effects occur?

DOI: 10.54254/2753-7064/2025.HT30156

This paper addresses both questions. First, it reviews the theoretical and methodological development of LR research, analyzing shifts from polarized debates on determinism versus universalism toward more conditional models. Second, it examines the most recent evidence from behavioral Verbal Interference Paradigms (VIP) and neural Event-Related Potentials (ERP) studies to highlight LR effects on thought versus behavior. Behavior studies show whether cognitive differences are strong and persistent enough to manifest without online verbal support, while ERP evidence reveals earlier, subtler modulations that may never cross behavioral thresholds. The contrast provides a conditional perspective and resolves apparent contradictions between "disappearing" and "hidden" effects.

2. Does language influence thought

2.1. Theoretical critiques and null evidence

The most longstanding and controversial contention concerns its theoretical falsifiability. Two lines of critics dominate the debate. The first targets the assumed causal relationship between language and thought. Critics often construct a reductive version of Whorfianism as equating language with thought. For instance, Pinker argues that humans think in an innate mental code of "mentalese," claiming that LR confuses the medium of expression with the substance of thought [12]. He further contends by stating that language cannot change basic physiological processes, as remarked: "no matter how influential language may be, it would seem preposterous to a physiologist that it could reach down into the retina and rewire the ganglion cells" [12]. From his view, language is a superficial layer on universal cognition and only the changes to the physiological architecture could alter cognition. These criticisms reflect the scholarly concern that LR may overstate language's influence of distorting empirical findings, such as the famous Eskimo Snow case. However, such critiques often oversimplify the original formulation of the hypothesis and risk undermining the scholarly value of this field. Hence, many scholars regard this line of critics as a straw man. In fact, Whorf neither proposed a formalized LR theory, nor equated language with thought, let alone claimed that cognition is determined by language. As Lucy notes, deterministic readings misrepresent Whorf's intent [13]. His notion of "worldview" was admittedly vague and lacks operational definition, which explains why modern neo-Whorfianism avoids strong determinism and adopts more cautious wording of "modulation" [14].

Moreover, the assumptions underlying Pinker's idea has been increasingly reassessed in the past decade, in which it suggests that the mind contains a core representation of "mental dictionary" or "encyclopedia" of stable, unchanging concepts. As commented by Casasanto, the static notion of core representation is fundamentally at odds with current understanding of the brain [15]. Rather than storing fixed ideas, the brain continually reshapes thoughts, whether they are concepts, percepts, feelings, or word meanings, which are instantiated in neural activity. As a result, scholars have advocated that the more important question is not the personal views of Sapir and Wholf, but rather the empirical relationship between language and thought. Like what Kay & Kempton wrote: "What either Sapir or Whorf actually believed on this topic is of course impossible to know, especially since the writings of both men are open to such varied interpretations. The question of what these two scholars thought, although interesting, is after all less important than the issue of what is the case" [16].

A second critique targets LR's inconsistent evidence. Scholars such as Pullum try to persuade readers that LR is too flexible to be empirically falsified [17]. In other words, it accommodates both confirming and disconfirming evidence, risking the status of a post-hoc explanation in pseudo-

DOI: 10.54254/2753-7064/2025.HT30156

science rather than a predictive theory. Classic counterexamples include Berlin and Kay's universal color term and Heider's studies with the Dani, whose limited lexicon did not impair recognition of focal hues [18, 10]. Other studies attribute apparent language effects to contextual factors. For instance, Li and Gleitman demonstrated that spatial reasoning strategies previously attributed to language in Brown and Levinson's work could instead arise from perceptually induced by landmark cues [11]. Similarly, Chen and O'Seaghdha suggested that reading and writing directions seems to have great impact in terms of Boroditsky's arguments on metaphorical thinking in English and Mandarin speakers' horizontal and vertical cognition of time [19]. Yet this framing treats LR as a binary claim: either language determines thought or it does not. In fact, apparent inconsistencies are expected, since "language" is not a monolithic variable but a complex, heterogeneous system varying typologically, functionally, and across populations. Null findings alone cannot establish theoretical invalidity and absence of evidence is not evidence of absence. Therefore, variability does not suggest theoretical weakness but reflects the conditional nature of linguistic modulation.

In addition, these criticisms raise a meta-scientific issue in contemporary psychology: the tendency to test hypotheses using null-hypothesis significance testing (NHST) before establishing a solid derivation chain linking theory, valid measurement, and causal assumptions. The apparent inconsistencies in LR evidence may reflect premature formal testing rather than theoretical incoherence. As Scheel et al. argue, non-confirmatory research is a prerequisite for interpretable hypothesis testing [20]. Therefore, the contested debates around LR exemplify a stage of theory development, which has motivated a series of experimental paradigm shifts in the following section.

2.2. Methodological development: circularity and causality

Previous debates lead to methodological evolution. Early studies often followed a language-centered approach: they began with a particular grammatical or lexical feature and predicted corresponding cognitive differences. This begins with Whorf's comparison of temporal encoding in Hopi versus English, and subsequent research on numerical cognition [1, 21, 22]. However, such designs risk circularity: if an experiment embeds the language in the procedure, any observed effect might simply reflect that bias [23].

To address this issue, researchers shifted to a behavior-centered approach, which infers cognition from cross-cultural behavioral differences that are assumed to be shaped by language (e.g., counterfactual reasoning [24]) and extended to ethnographic and socio-economic research (e.g., accident rates, savings behavior, and pro-environmental decision-making [25-27]). These studies offer broad applied insights, but causal inference from behavior to language often relies on interpretive assumptions about cognitive mechanisms, risking confounds with education, culture, or other sociological factors.

The most recent domain-centered approach attempts to isolate cognition from language in experiments using non-verbal tasks and controlled language exposure. The Whorfian Psychophysics Paradigm (WPP) is a significant contribution that applies psychophysical techniques by examining how precisely physical stimuli are perceived, discriminated, and reproduced [15]. In this paradigm, participants first completed a non-linguistic psychophysical task (e.g., estimating duration, distance, or pitch) to address circularity, and then learned new linguistic mappings to assess conceptual change to establish causality. For example, the training of Dutch speakers on Farsi pitch metaphors subsequently shifted their pitch judgments from native language modulation [28]. These studies establish the causal impact of language on cognition and shift the focus from whether language shapes cognition to more conditional models: frameworks that specify when and under what contextual conditions LR manifest.

DOI: 10.54254/2753-7064/2025.HT30156

3. Magnitude and malleability

3.1. Domain specificity

Conceptual domain is an important condition. In 2001, Boroditsky proposed that concepts in concrete domains (e.g., color) can be learned by perceptual interaction with the environment and before language acquisition. By contrast, concepts in abstract domains (e.g., time) lack such immediate grounding and are predicted to be more dependent on linguistic scaffolding [2]. This contradiction is supported by both theoretical and developmental accounts. Dual Coding Theory and Vygotsky's concept of linguistic scaffolding predict that concrete concepts possess both visual and verbal codes, whereas abstract concepts rely primarily on verbal codes for learning and representation [29, 30]. Furthermore, Gillette and colleagues reported that children acquire object-based, concrete referents more rapidly and reliably than abstract relational terms [31].

Neurocognitive findings from embodied semantics also support this distinction. Pulvermüller found that action verbs activate motor regions within approximately 200 ms, linking semantic activation directly with sensorimotor systems [32]. Most importantly, a meta-analysis of 72 neuroimaging studies shows that concrete and abstract concepts vary both in the information-processing modalities and cognitive control demands. Their results conclude that abstract concepts preferentially recruited networks associated with social cognition, language, and semantic control, whereas concrete concepts predominantly activated action-related regions [33]. These findings suggest linguistic modulation tends to manifest in abstract domains.

3.2. Temporal persistence

3.2.1. Online processing

Another extensively researched axis is timescale. Slobin's Thinking-for-Speaking theory (TFS) was one of the earliest frameworks: utterances guide real-time attention and encoding without permanently restructuring thought [34]. Verbal Interference Paradigms (VIP) have provided key empirical support for this online-processing perspective [5, 7, 35]. For example, Russian speakers' advantage in discriminating shades of blue disappears under concurrent verbal tasks, implying reliance on active inner speech or lexical access [35]. Based on VIP, Label-Feedback Hypothesis offers a mechanistic account for this online processing perspective [36]. In Lupyan's model (see Figure 1.), each time a label is used, it sends top-down feedback to perception, while sensory input reactivates the label, forming a recurrent loop that shapes perceptual processing. If inner speech is blocked by verbal suppression, this enhancement disappears. In his model, perception is treated as a hybrid of bottom-up sensory input and top-down linguistic input.

DOI: 10.54254/2753-7064/2025.HT30156

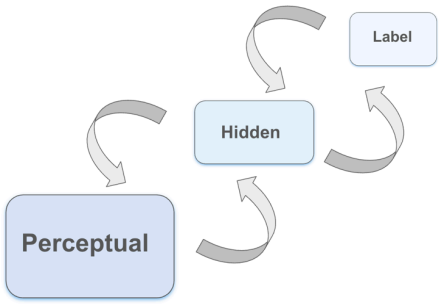

Figure 1. A schematic of a neural network architecture for exploring on-line effects of labels on perceptual representations. Adapted from [36]

On the one hand, subsequent VIP studies have replicated such patterns across domains. Scholars increasingly regard LR effects as online and transient, emerging when linguistic resources are available and fading when they are blocked. On the other hand, with emerging offline evidence, scholars continue to develop sophisticated models beyond online processing. Early accounts often relied on stage-based models, but recent work emphasizes more on dynamic frameworks that allow flexible interactions between language and cognition across multiple timescale.

3.2.2. Before/With/After framework

One of the early attempts is Wolff and Holmes's Before/With/After framework. As illustrated in Figure 2., the model provides a clear classification for linguistic modulation occurring (a) before speech, during utterance preparation; (b) with speech, during real-time linguistic engagement; or (c) after speech, affecting memory, conceptual organization, or behavior over longer timescales [37].

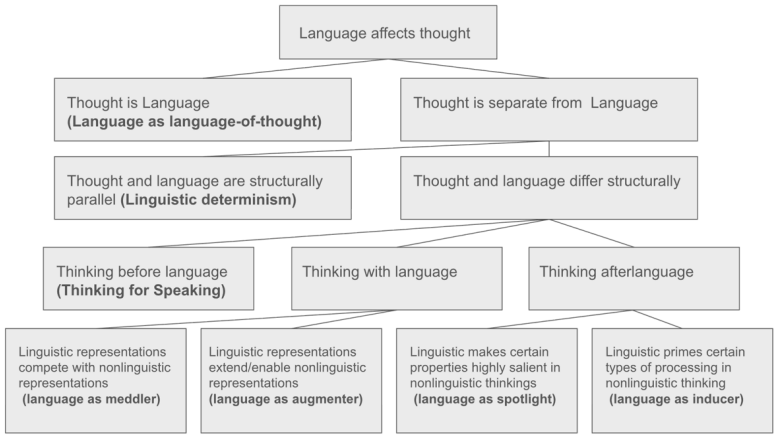

Figure 2. Classes and subclasses of hypotheses on how language might affect thought. Adapted from [37]

However, assigning specific mechanisms neatly to these categories remains debatable. For example, TFS is classified as a "before" effect, yet it features more consistent with an online process because it occurs during utterance planning and is contingent on linguistic formulation. Similarly, the placement of color categorization effects (e.g., Winawer et al. [35]) as "with-speech" or online "meddler" effects overlooks evidence from ERP studies (e.g., Thierry et al. [38]). ERP evidence shows that color perception is influenced by long-term linguistic accumulation at pre-attentive

DOI: 10.54254/2753-7064/2025.HT30156

stages of early visual processing, indicating automatic effects rather than purely online ones. As a result, many researchers are shifting toward models that treat language effects as recursive and multi-timescale.

3.2.3. Ad-Hoc Cognition

Furthermore, Casasanto and Lupyan's Ad Hoc Cognition framework proposes that concepts are not fixed entities, but momentary assemblies built from language, perception, and context [39]. They outline three timescales of modulation: (a) Activation dynamics (millisecond-by-millisecond: online effect); (b) Local context effects (minutes to hours: short-term differences as tasks or primes change one's state); and (c) Experiential relativity (days to lifetimes: long-term developmental differences across people or cultures). In their experiment, English and Greek speakers differ in their default spatial metaphors mapping for time as a long-term modulation of experiential relativity, yet brief training can temporarily reverse the mapping direction under local context effects [40]. Their mechanism reframes LR as a dynamic process that flexibly emerges under different temporal and contextual conditions.

4. Distinction between behavior and thought

4.1. Behavior methods: Verbal Interference Paradigms (VIP)

In principle, Verbal Interference Paradigms (VIP) temporarily block verbal working memory to test whether cognitive performance still reflects linguistic modulation. This approach is crucial for distinguishing effects that rely on activation dynamics from experiential relativity, since participants may sub-vocally "speak to themselves" during nonverbal tasks.

Recent reviews have identified 4 commonly used manipulations: syllable repetition, verbal short-term memory tasks, shadowing, and verbal judgment tasks, and suggests that covert speech primarily supports simple categorization, memory, and task switching of nameable stimuli, but not visuospatial or reasoning tasks [41]. This pattern indicates that online verbal processes tend to facilitate tasks requiring symbolic labeling rather than perceptual integration. Accordingly, experimental designs must ensure that observed effects reflect disrupted speech rather than cognitive load. As spatial tasks rely minimally on verbal mediation, they are commonly used as comparison tasks in VIP. Empirically, many classic LR effects vanish under VIP, as in color discrimination [35], exact number [4, 5], and motion events [7]. Therefore, researchers increasingly interpret LR as online and transient phenomena until recent ERP evidence emerges.

4.2. Neurolinguistic methods: Event-Related Potentials (ERP)

Derived from electroencephalography (EEG), ERP provides a complementary perspective by bypassing behavior altogether. Behavioral measures like accuracy rate (ACC) and reaction time (RT) represent the end product of complex cognitive processes and are influenced by multiple factors. On the contrary, ERP captures millisecond-level voltage changes in brain activity time-locked to specific events without the need for overt responses. Different ERP components index different linguistic processes. For example, N100 for early visual and attentional processing, N400 for semantic integration, and P600 for syntactic reanalysis [42-44].

ERP studies show that effects previously attributed to online linguistic processing in VIP can arise independently of overt language use. For instance, Thierry et al. reported that Greek speakers, who lexically differentiate light and dark blue, show stronger early responses (~100 ms) to blue-

DOI: 10.54254/2753-7064/2025.HT30156

color changes than English speakers [38]. Most importantly, the effect is thought to precede online verbal processing, which means that linguistic modulation in early color perception can occur unconsciously [38]. At the same time, Flecken et al. found that German speakers, whose grammar marks path of motion, showed a distinct P3-wave response to changes in motion endpoints in the 350–700 ms time window, whereas English speakers show no difference [45]. Although this component occurs later than the N100 window and does not directly rule out covert verbalization, the researchers argue that there were no differences in participants' behavioral similarity judgement results (both ACC and RT), suggesting that the stimuli were highly abstract and difficult to verbalize spontaneously. Moreover, the P3 effects did not align with the timing or polarity of classic semantic processing components of N400, further supporting the conclusion that these effects are automatic rather than strategy based [45]. The neural signatures suggest offline LR effects could occur beyond strategic behavior.

4.3. Distinction between behavior and thought

However, a tension arises when comparing these ERP findings to behavioral measurements. In both color and motion domain, VIP studies show that LR effects depend on online labeling [7, 35]. By contrast, ERP studies reveal that cognitive distinctions persist in these domains without overt language use [38, 45]. The key lies in that neuro-cognitive activation, and overt behavioral outcomes may reflect different levels of processing. Neural activation may capture subtle linguistic biases that fail to reach conscious decision-making.

This argument is supported by studies combining ERP and behavior measurements. In the aforementioned study conducted by Flecken and colleagues, they performed a follow-up experiment using a motion-matching task [45]. Unlike distinct ERP results, the behavioral test indicates that such effects did not translate into overt behavior outcomes. Similarly, Liu et al. conducted both ERP study and behavioral judgement during color discrimination tasks [46]. Although behavioral performance showed no group differences, ERP data revealed stronger N2 and delayed P3 responses between 260 and 310 ms post-stimulus. The difference only appears in the right visual field, which is processed by the language-dominant hemisphere. This pattern suggests that language can modulate cognition without affecting behavior. Compared with Thierry et al., Liu et al. found later ERP effects and no modulation of early visual components (e.g., N1, P1) [38, 45]. They attributed this distinction to task demands: Thierry's oddball design required only change detection, whereas their study required explicit comparison of two simultaneously presented colors, likely engaging higher-level prefrontal processing.

These findings highlight that the absence of behavioral effects does not imply the absence of cognitive or neural influence. EEG can detect subtle, automatic, and pre-verbal modulations that may never reach behavior threshold. Meanwhile, VIP functions as a decision boundary task by identifying whether LR effects are strong enough to guide deliberate action. In conclusion, the apparent contradiction between the two paradigms should not be interpreted as evidence against LR but reflects different levels of cognitive processing.

5. Discussion: a conditional perspective

Current evidence supports a conditional view toward the relationship between language and cognition. Language influences cognition across domains, but the observed persistence varies by domain and method. Behavioral paradigms suggest most linguistic modulations require concurrent language involvement, whereas ERP studies show earlier and implicit modulations, implying that

DOI: 10.54254/2753-7064/2025.HT30156

language can shape perceptual and conceptual processing at preconscious levels. In particular, VIP findings are critical because they potentially separate effects dependent on active inner speech from those rooted in deeper representational restructuring.

Moving beyond a monolithic notion of "language," these findings demonstrate a continuum of influence from lexical labels to grammatical and metaphorical structures (see Figure 3.). At the lexical level, words and morphemes are symbolic form–meaning pairings [47, 48]. From a constructionist perspective, lexical labels simply map words onto perceptual categories (e.g., blue to a region of color space, two to a discrete quantity) without reconceptualizing internal structure. Therefore, labeling could help with sharpening distinction and enhancing memory, but are arbitrary, unstructured, and tend to rely on online verbal modulation. For instance, LR advantages in the color domain often disappear under verbal interference [35], and numerical cognition show similar online dependency [4, 5].

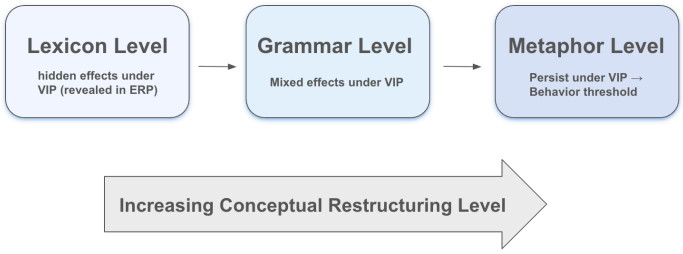

Figure 3. Hierarchical framework of linguistic influence revealed by VIP and ERP paradigms

At the grammatical level, linguistic structures encode relations and perspectives (e.g., aspect, voice, thematic roles) in obligatory and structural ways, showing a more stable influence on conceptualization than lexical labels. The mixed VIP evidence supports this interpretation: although motion-event encoding depends on overt language and disappears under verbal interference [7], grammatical gender effects persist [8]. This persistence may arise from the conceptual projection that grammatical gender shares with metaphor, in which abstract qualities (e.g., masculinity or femininity) are extended to inanimate objects. However, unlike metaphor, grammatical gender is a morphosyntactic system, which is form-based, automatic, and deeply embedded in sentence structure. These findings locate grammar at an intermediate level of LR, which is not entirely independent of utterance.

At the metaphorical level, language extends beyond literal reference, usually using concrete, embodied schemas to conceptualize abstract domains, which is a central premise of Conceptual Metaphor Theory [49]. This claim has been empirically supported by behavioral, developmental, and neuro-scientific research. In this view, conceptual mappings reorganize conceptual structure and are not tied to surface form and overt language use, enabling operating below the level of conscious access and guiding cognitive organization even in non-verbal tasks. For instance, studies focused on metaphorical mapping have shown that spatial metaphors for pitch ("thick/thin" vs. "high/low") and time ("long/short" vs. "small/big") still influence nonverbal judgments under verbal interference [28, 40]. The persistence under verbal suppression marks the deepest level of LR effect, in which language shapes cognition implicitly and representationally.

DOI: 10.54254/2753-7064/2025.HT30156

6. Conclusion

The present review reconciles seemingly conflicting evidence from VIP and ERP studies by offering a distinct perspective on behavioral outcomes and underlying cognitive processes. Based on this proposal, ERP and VIP evidence support a hierarchical view of LR. Different levels of linguistic devices have different levels of modulation that differ in their degree of automaticity, representational depth, and temporal persistence. VIP identifies whether the effects depend on active linguistic support, whereas ERP uncovers implicit, preconscious biases in neural processing.

Future work should more systematically integrate behavioral and neural measures within the same experimental frameworks, while carefully manipulating task demands, conceptual domain, and temporal resolution. Such designs will allow researchers to map more precisely the conditions under which linguistic influences are “disappearing” at the behavioral level yet remain “hidden” in neurocognitive activity. The central question in this field is no longer to ask whether language modulates thought, but to investigate when, how, and at what level such modulation occurs.

References

[1] Whorf, B. L. (1956). Language, Thought, and Reality: Selected Writings. Cambridge, MA: The MIT Press.
[2] Boroditsky, L. (2001). Does Language Shape Thought? Mandarin and English Speakers’ Conceptions of Time. Cognitive Psychology 43, 1–22. doi: 10.1006/cogp.2001.0748
[3] Casasanto, D., and Boroditsky, L. (2008). Time in the mind: Using Space to Think About Time. Cognition 106, 579–593. doi: 10.1016/j.cognition.2007.03.004
[4] Gordon, P. (2004). Numerical Cognition without Words: Evidence from Amazonia. Science 306, 496–499. doi: 10.1126/science.1094492
[5] Frank, M. C., Fedorenko, E., Lai, P., Saxe, R., and Gibson, E. (2012). Verbal Interference Suppresses Exact Numerical Representation. Cogn Psychol 64, 74–92. doi: 10.1016/j.cogpsych.2011.10.004
[6] Oh, K.-J. (2003). Language, Cognition, and Development: Motion Events in English and Korean. [PhD dissertation]. [Berkeley (CA)]: University of California, Berkeley.
[7] Trueswell, J. C., and Papafragou, A. (2010). Perceiving and Remembering Events Cross-Linguistically: Evidence from Dual-task Paradigms. Journal of Memory and Language 63, 64–82. doi: 10.1016/j.jml.2010.02.006
[8] Boroditsky, L., Scmidt, L. A., and Phillips, W. (2003). “Sex, Syntax and Semantics, ” in Language in Mind: Advances in the Study of Language and Thought, (Cambridge, MA: The MIT Press), 61–79. doi: 10.7551/mitpress/4117.001.0001
[9] Boutonnet, B., Athanasopoulos, P., and Thierry, G. (2012). Unconscious Effects of Grammatical Gender during Object Categorisation. Brain Res 1479, 72–79. doi: 10.1016/j.brainres.2012.08.044
[10] Heider, E. R., and Olivier, D. C. (1972). The Structure of the Color Space in Naming and Memory for Two Languages. Cognitive Psychology 3, 337–354. doi: 10.1016/0010-0285(72)90011-4
[11] Li, P., and Gleitman, L. (2002). Turning the Tables: Language and Spatial reasoning. Cognition 83, 265–294. doi: 10.1016/S0010-0277(02)00009-4
[12] Pinker, S. (1999). The Language Instinct: How the Mind Creates Language., Original work published 1994. New York, NY: HarperPerennial.
[13] Gumperz, J. J., and Levinson, S. C. (1996). Rethinking Linguistic Relativity. Cambridge University Press.
[14] Bohnemeyer, J. (2020). “Linguistic Relativity: From Whorf to Now, ” in The Wiley Blackwell Companion to Semantics, eds. D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, and T. Zimmermann (Wiley), 1–33. doi: 10.1002/9781118788516.sem013
[15] Casasanto, D. (2016). Linguistic relativity. In N. Riemer (ed.), The Routledge Handbook of Semantics (pp. 158–174). New York, NY: Routledge.
[16] Kay, P., and Kempton, W. (1984). What is the Sapir-Whorf hypothesis? American Anthropologist 86, 65–79. doi: 10.1525/aa.1984.86.1.02a00050
[17] Pullum, G. K. (1991). The Great Eskimo Vocabulary Hoax and Other Irreverent Essays on the Study of Language. Chicago, IL: University of Chicago Press. Available at: https://press.uchicago.edu/ucp/books/book/chicago/G/bo3684610.html (Accessed October 14, 2025).

DOI: 10.54254/2753-7064/2025.HT30156

[18] Berlin, B., and Kay, P. (1991). Basic Color Terms: Their Universality and Evolution. Original work published 1969. Berkeley, CA: University of California Press.
[19] Chen, J. Y., and O'Seaghdha, P. G. (2013). Do Mandarin and English Speakers Think about Time Differently? Review of Existing Evidence and Some New Data. Journal of Chinese Linguistics 41, 338–358.
[20] Scheel, A. M., Tiokhin, L., Isager, P. M., and Lakens, D. (2021). Why Hypothesis Testers Should Spend Less Time Testing Hypotheses. Perspect Psychol Sci 16, 744–755. doi: 10.1177/1745691620966795
[21] Lucy, J. A. (1992). Language Diversity and Thought: A Reformulation of the Linguistic relativity hypothesis. New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511620843
[22] Miller, K. F., Smith, C. M., Zhu, J., and Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: The role of number-naming systems. Psychological Science 6, 56–60. doi: 10.1111/j.1467-9280.1995.tb00305.x
[23] Pinker, S. (2007). The stuff of Thought: Language as a Window into Human Nature. NY: Viking.
[24] Bloom, A. H. (2014). The Linguistic Shaping of Thought: A Study in the Impact of Language on Thinking in China and the West. Original work published 1981. New York, NY: Psychology Press. doi: 10.4324/9781315802893
[25] Salminen, S., and Johansson, A. (2000). Occupational accidents of Finnish- and Swedish-speaking Workers in Finland: a Mental Model View. Int J Occup Saf Ergon 6, 293–306. doi: 10.1080/10803548.2000.11076456
[26] Chen, M. K. (2013). The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets. American Economic Review 103, 690–731. doi: 10.1257/aer.103.2.690
[27] Mavisakalyan, A., Tarverdi, Y., and Weber, C. (2018). Talking in the Present, Caring for the Future: Language and Environment. Journal of Comparative Economics 46, 1370–1387.
[28] Dolscheid, S., Shayan, S., Majid, A., and Casasanto, D. (2013). The Thickness of Musical Pitch: Psychophysical Evidence for Linguistic Relativity. Psychol Sci 24, 613–621. doi: 10.1177/0956797612457374
[29] Paivio, A. (1991). Dual Coding Theory: Retrospect and Current Status. Canadian Journal of Psychology / Revue canadienne de psychologie 45, 255–287. doi: 10.1037/h0084295
[30] Vygotsky, L. S. (1994). Thought and Language. Cambridge, MA, US.: The MIT Press.
[31] Gillette, J., Gleitman, H., Gleitman, L., and Lederer, A. (1999). Human Simulations of Vocabulary Learning. Cognition 73, 135–176. doi: 10.1016/S0010-0277(99)00036-0
[32] Pulvermüller, F. (2005). Brain Mechanisms Linking Language and Action. Nat Rev Neurosci 6, 576–582. doi: 10.1038/nrn1706
[33] Hoffman, P., and Bair, M. (2025). How do Brain regions Specialised for Concrete and Abstract Concepts align with Functional Brain Networks?: A Neuroimaging Meta-Analysis. Neuroscience & Biobehavioral Reviews 174, 106214. doi: 10.1016/j.neubiorev.2025.106214
[34] Slobin, D. I. (1996). From "Thought and Language" to "Thinking for Speaking.". New York, NY: Cambridge University Press.
[35] Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., and Boroditsky, L. (2007). Russian Blues Reveal Effects of Language on Color Discrimination. Proc Natl Acad Sci U S A 104, 7780–7785. doi: 10.1073/pnas.0701644104
[36] Lupyan, G. (2012). Linguistically Modulated Perception and Cognition: The Label-Feedback Hypothesis. Frontiers in Psychology 3. doi: 10.3389/fpsyg.2012.00054
[37] Wolff, P., & Holmes, K. J. (2011). Linguistic Relativity. WIREs Cognitive Science, 2(3), 253–265. https: //doi.org/10.1002/wcs.104
[38] Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., and Kuipers, J.-R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. Proc Natl Acad Sci U S A 106, 4567–4570. doi: 10.1073/pnas.0811155106
[39] Casasanto, D., and Lupyan, G. (2015). "All Concepts Are Ad Hoc Concepts, " in The Conceptual Mind, eds. E. Margolis and S. Laurence (Cambridge, MA, US: The MIT Press), 543–566. doi: 10.7551/mitpress/9383.003.0031
[40] Casasanto, D., Fotakopoulou, O., Pita, R., and Boroditsky, L. (2025). The Long and Small of it: Language Shapes Duration Estimation in Speakers of English and Greek. Language and Cognition. (in press).
[41] Nedergaard, J. S. K., Wallentin, M., and Lupyan, G. (2023). Verbal interference Paradigms: A Systematic Review Investigating the Role of Language in Cognition. Psychon Bull Rev 30, 464–488. doi: 10.3758/s13423-022-02144-7
[42] Mangun, G. R., Buonocore, M. H., Girelli, M., and Jha, A. P. (1998). ERP and fMRI Measures of Visual Spatial Selective Attention. Hum Brain Mapp 6, 383–389. doi: 10.1002/(SICI)1097-0193(1998)6: 5/6& %23x0003c; 383: : AID-HBM10& %23x0003e; 3.0.CO; 2-Z
[43] Kutas, M., and Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-related Brain Potential (ERP). Annu Rev Psychol 62, 621–647. doi:

DOI: 10.54254/2753-7064/2025.HT30156

10.1146/annurev.psych.093008.131123
[44] Kaan, E., Harris, A., Gibson, E., and Holcomb, P. (2000). The P600 as an Index of Syntactic Integration Difficulty. Language and Cognitive Processes 15, 159–201. doi: 10.1080/016909600386084
[45] Flecken, M., Athanasopoulos, P., Kuipers, J. R., and Thierry, G. (2015). On the road to Somewhere: Brain Potentials Reflect Language Effects on Motion Event Perception. Cognition 141, 41–51. doi: 10.1016/j.cognition.2015.04.006
[46] Liu, Q., Li, H., Campos, J. L., Teeter, C., Tao, W., Zhang, Q., et al. (2010). Language Suppression Effects on the Categorical Perception of Colour as Evidenced through ERPs. Biological Psychology 85, 45–52. doi: 10.1016/j.biopsycho.2010.05.001
[47] Saussure, F. (1916). Course in General Linguistics. London: Duckworth.
[48] Goldberg, A. E. (2013). Constructionist Approaches., eds. T. Hoffmann and G. Trousdale. Oxford University Press. doi: 10.1093/oxfordhb/9780195396683.013.0002
[49] Lakoff, G., and Johnson, M. (2003). Metaphors We Live By. Chicago, IL: University of Chicago Press.